Multivariate Age-Period-Cohort Models

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Andrea Riebler

aus

Deutschland

Promotionskomitee

Prof. Dr. Leonhard Held (Vorsitz)
Prof. Dr. Andrew Barbour
Prof. Dr. Peter Bühlmann (ETH Zürich)

Zürich, 2010

Preface

I owe thanks to many people who supported and encouraged me during the last three years while I was working on my PhD thesis.

First of all I would like to thank Leonhard Held for giving me the possibility to write this thesis and being an excellent supervisor. I am grateful for his continuous support and all the fruitful discussions. I am also very thankful for giving me the opportunity to spend two research visits in Norway and to attend several national and international conferences which broadened my mind.

I would also like to thank all my colleagues of the Biostatistics Unit for building such an inspiring and supportive working atmosphere. My special thanks goes to Michaela Paul, for sharing an office with me over the last three years and many productive discussions. I thank my coauthor Matthias Bopp for helpful discussions about suicide mortality in Switzerland. I am grateful to Sinikka Kohler and Hanspeter Jauss for all administrative help and my thanks also go to all those who provided useful comments during the past three years and proofread parts of this thesis.

Moreover, I would like to thank Håvard Rue and his group for many interesting discussions and the hospitality during my two research stays at the Norwegian University of Science and Technology in Trondheim.

I express my thanks to Sylvia Frühwirth-Schnatter who kindly accepted to review my thesis. Furthermore, I am grateful to Andrew Barbour and Peter Bühlmann for being part of my dissertation committee.

Finally, I would like to thank my mother for supporting me and being always there. My very special thanks go to Christian for his constant presence, encouragement and infinite patience.

Financial support by the Swiss National Science Foundation and partly by the Research Council of Norway is gratefully acknowledged.

Zurich, May 2010

Andrea Riebler

Zusammenfassung

Alters-Perioden-Kohorten (APK) Modelle werden zur Analyse von altersspezifischen Krankheits- oder Sterberaten verwendet, die über mehrere Kalenderperioden erfasst wurden. Die Daten werden im Hinblick auf drei Zeitskalen analysiert: Alter, Periode (Infektions- oder Todeszeitpunkt) und Kohorte (Geburtszeitpunkt). Häufig liegen mehrere solcher Datentabellen vor, da die Daten anhand einer weiteren Stratifizierungsvariable archiviert wurden, so dass für jede Ausprägung (Schicht) dieser Variable ein Datensatz existiert. Beispielsweise könnten Raten für Männer und Frauen oder für mehrere geographische Regionen verfügbar sein. Jeder einzelne dieser Datensätze könnte separat mit einem univariaten APK Modell analysiert werden. Jedoch könnte es auf Grund von ähnlichen Risikofaktoren von Vorteil sein, alle Datensätze gemeinsam zu analysieren, wobei manche Zeiteffektgruppen als gemeinsam über die Schichten behandelt werden können. Diese Dissertation hat zum Ziel, die Methodik für die statistische Inferenz von multivariaten APK Modellen weiterzuentwickeln.

Zunächst zeigen wir, dass Differenzen von schichtspezifischen Zeiteffekten in multivariaten APK Modellen identifizierbar sind, so dass das für univariate APK Modelle bekannte Identifizierbarkeitsproblem vermieden wird. Wir entwickeln ein multivariates Bayesianisches Modell basierend auf Glättungsprioris, um heterogene Zeittrends zu analysieren. Dieser Ansatz repräsentiert eine attraktive Alternative zu Maximum Likelihood (ML) basierten Ansätzen, wenn Altersgruppen und Perioden für die gleichen Zeitintervallbreiten erfasst wurden, und vermeidet die Artefakte, z.B. unechte zyklische Muster, die im Fall von ungleichen Zeitintervallbreiten auftreten.

Anschliessend präsentieren wir einen bedingten Ansatz zur Inferenz in multivariaten APK Modellen. Im Gegensatz zum unbedingten Ansatz, der viele Nuisance-Parameter enthält, modelliert der bedingte Ansatz direkt die relevanten Parameter, nämlich die Differenzen von schichtspezifischen Zeiteffekten. Darüber hinaus erweitern wir diesen Ansatz, um Datensätze mit multiplen Stratifizierungsfaktoren zu analysieren. Die Schätzung mittels ML basiert auf Standardsoftware für multinomiale logistische Regression. Die Verwendung von kubischen Glättungssplines wird vorgeschlagen, um unechte zyklische Muster im Fall von unterschiedlich weiten Zeitintervallen von Alter und Periode zu vermeiden.

Zuletzt schlagen wir die Verwendung von korrelierten Glättungsprioris und korrelierten Überdispersionsparametern vor, um die eventuelle Abhängigkeit zwischen multiplen Datensätzen zu erfassen. Mittels Fallstudien zeigen wir, dass korrelierte multivariate APK Modelle nützlich sind, um die Präzision von geschätzten relativen Risiken zu verbessern und fehlende Daten zu extrapolieren. Wir implementieren die Methodik mit Markov-Ketten-Monte-Carlo (MCMC) und der kürzlich vorgeschlagenen integrierten genesteten Laplace Approximation (INLA). Mit INLA ist es möglich, eine Vielzahl an weiteren latenten Gauß Modellen zu korrelieren, z.B. bedingte autoregressive Modelle oder saisonale Modelle.

In einer Anwendung auf Schweizer Selbstmorddaten von 1950 bis 2007 analysieren wir geschlechtsspezifische Unterschiede mittels einfachen und korrelierten multivariaten APK Modellen. Die Ergebnisse deuten an, dass Männer das ca. dreifache Risiko von Frauen haben, Selbstmord zu begehen. Ältere Männer und $15-24$ Jährige sind besonders gefährdet. Des Weiteren benutzen wir uni- und multivariate APK Modelle, um zu untersuchen, ob erklärende Variablen mit Bezug zu Familienintegration geschlechtsspezifische Unterschiede erklären können.

Abstract

Age-period-cohort (APC) models are used to analyse age-specific disease or mortality rates provided for several periods in time with respect to three time scales: age, period (calendar period during which the incidence or mortality rates were observed) and cohort (time of birth). Frequently, several sets of such age-specific rates are observed because data were recorded according to one further stratification variable resulting in one set of rates for each stratum of this variable. For example, rates might be available for males and females or for several geographical regions. Each set of rates could be analysed separately by means of an univariate APC model. However, because of similar relevant risk factors it might be beneficial to analyse all sets of rates jointly treating some sets of time effects as common across strata. Multivariate APC models share sets of time effects, for example the age effects, while the remaining parameters can be different. This dissertation aims at improving the methodology for statistical inference in multivariate APC models.

We first show that differences of stratum-specific time effects in multivariate APC models are identifiable, so that the well known identifiability problem for univariate APC models is avoided. We develop a multivariate Bayesian APC model based on smoothing priors to analyse heterogeneous time trends. This approach represents an attractive alternative to maximum likelihood (ML) based approaches when age groups and periods are given for the same time-interval widths and avoids the artefacts, e.g. artificial cyclical patterns, which occur in the case of unequal time-interval widths.

Subsequently, we present a conditional approach for inference in multivariate APC models. In contrast to the unconditional approach which includes many nuisance parameters, the conditional approach directly models the parameters of interest, namely the differences of stratum-specific time effects. Furthermore, we extend this approach to analyse datasets with multiple stratification factors. ML estimation is performed using software for multinomial logistic regression. The use of cubic smoothing splines is proposed to avoid artificial cyclical patterns in the case of unequally spaced time-intervals of age and period.

Finally, we propose the use of correlated smoothing priors and correlated overdispersion parameters to capture the potential dependence present between multiple health outcomes. By means of case studies we demonstrate that correlated multivariate APC models are useful to improve the precision of relative risk estimates and to extrapolate missing data. We implement the methodology using Markov chain Monte Carlo (MCMC) and the recently proposed integrated nested Laplace approximation (INLA). With INLA it is possible to correlate a wide range of other latent Gaussian models, e.g. conditionally autoregressive models or seasonal models.

In an application to Swiss suicide data from 1950 to 2007, we analyse gender-specific differences using both ordinary and correlated multivariate Bayesian APC models. Results indicate that males have approximately three times the risk of committing suicide as women. Elderly men and those between $15 - 24$ are especially at risk. Furthermore, we use univariate and multivariate APC models to investigate whether explanatory variables related to family integration can explain gender-specific suicide trends.

Thesis outline

Introduction

Paper I: **The analysis of heterogeneous time trends in multivariate age-period-cohort models**

Andrea Riebler & Leonhard Held

Paper published in *Biostatistics (2010), 11, 1, pp. 57–69.*

Paper II: **A conditional approach for inference in multivariate age-period-cohort models**

Leonhard Held & Andrea Riebler

Paper accepted for publication in *Statistical Methods in Medical Research.*

Paper III: **Correlated multivariate age-period-cohort models**

Andrea Riebler, Leonhard Held & Håvard Rue

Technical Report, University of Zurich.

Paper IV: **Suicide mortality in Switzerland: Gender-specific differences and the impact of family integration**

Andrea Riebler, Leonhard Held, Håvard Rue & Matthias Bopp

Technical Report, University of Zurich.

Appendix I: **Correlated GMRF priors for multivariate age-period-cohort models**

Andrea Riebler, Leonhard Held & Håvard Rue

Paper published in *Proceedings of the 25th International Workshop on Statistical Modelling. Glasgow, July 5-9, 2010, Adrian W. Bowman (editor), pp. 455–460.*

Appendix II: **Full conditional distributions**

Appendix III: **Program description**

Introduction: Age-period-cohort models

Age-period-cohort (APC) models are used to describe and predict time series of disease or mortality rates using three different time scales: age, period (calendar period during which the incidence or mortality rates were observed) and cohort (longitudinally observed group of people born within specific periods). In the following an introduction to the main aspects of APC analysis will be given.

1 The Lexis diagram

Most developed countries have national health care registers collecting data on mortality and disease rates. For new cases typically diagnosis, gender, age, date of diagnosis, etc. are recorded. To represent such tabulated records graphically the Lexis diagram, named after Wilhelm Lexis (Lexis, 1875), is frequently used (Keiding, 1990). The Lexis diagram is a two-dimensional coordinate system. Calendar time (period) is represented on the horizontal axis, while age is represented on the vertical axis. Both axes are divided into intervals which partition the diagram into squared compartments. Every compartment refers to a group of persons that belongs to a specific age group at a specific period. Figure 1 shows the Lexis diagram for an example. The

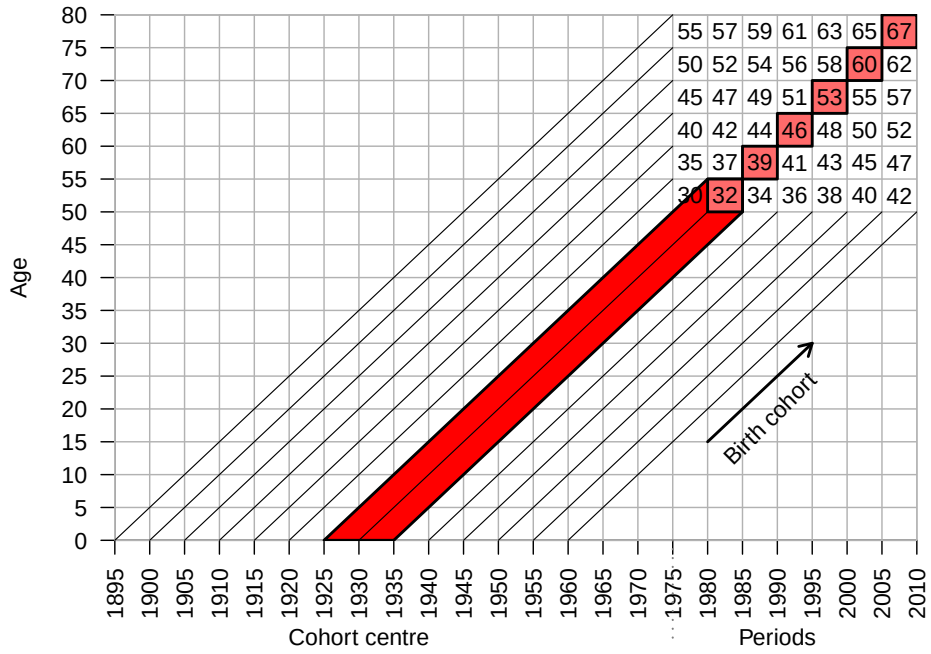

Figure 1: Lexis diagram for a toy example.

numbers of deaths or diseases (and analogously the number of persons under risk) stratified by age and period can be written into the corresponding compartments of the diagram. For instance, in Figure 1 there were 32 cases between 1980 and 1985 for persons aged between 50 and 55. A third time scale, namely the birth cohort, is clearly determined by the age group and calendar period. Drawing a line from the upper left corner and the lower right corner of a compartment to the horizontal axis with an angle of 45° determines the beginning and the end

of the birth cohort the compartment belongs to. For instance, in Figure 1 the red shaded area represents the data of the cohort beginning in 1925 and ending in 1935. As can be seen the cohorts overlap. If periods and age groups are defined on the same time grid drawing a line from the lower left corner of a data square to the abscissa with an angle of 45° leads to the centre of a cohort.

The three time scales age, period and birth cohort are often regarded to be linked to death caused by disease. However, Berzuini and Clayton (1994) emphasised that time itself does not cause disease events and that it is only a scale on which other factors operate. Time is a surrogate or proxy for other causal factors that are unknown or difficult to measure. For example, in chronic degenerative diseases, age is related to cumulative factors that increase the risk of contracting the disease. Factors that explain changes in rates across different age groups are called age effects. Factors that apply to all persons at a particular calendar time independent of their age are called period effects, for instance medical advances in treatment or environmental changes. Cohort effects seek to explain changes across different birth cohorts, for example generation-specific habits. To analyse registry data regarding the three time scales age, period and cohort, the use of APC models was proposed in the fields of demography, epidemiology and also sociology, see for example Fienberg and Mason (1979); Holford (1983); Clayton and Schifflers (1987b).

2 The univariate APC model

Let n_{ij} denote the number of persons at risk in age group i ($i = 1, \ldots, I$) during period j ($j = 1, \ldots, J$). We assume that the number of cases y_{ij} in age group i at period j follows a Poisson distribution with rate $n_{ij} \times \lambda_{ij}$ and that the likelihood for the whole data is given by the product of the corresponding Poisson terms. The classical APC model (Clayton and Schifflers, 1987b) decomposes the linear predictor $\log \lambda_{ij}$ additively into an overall level (intercept) μ and age effect θ_i, period effect φ_j and cohort effect ψ_k, so that

$$\eta_{ij} = \log \lambda_{ij} = \mu + \theta_i + \varphi_j + \psi_k.$$

Here $k = 1, \ldots, K$ refers to the birth cohort index and depends directly on the age index i and the period index j. In the case of equal interval widths for age group and period $k = (I - i) + j$ and $K = (I - 1) + J$. In the context of the Lexis diagram (Figure 1) index i represents the row index, index j the column index and k the diagonal bands from the lower left to the upper right.

However, if age group and period are defined on different time grids, a slightly different definition of k has to be used. Holford (1983) consider the case in which the period intervals are M times wider than the age group intervals. Holford (2006) analyse data with five-year intervals for age and three-year intervals for period with a total of $I = 11$ age groups and $J = 16$ periods. They indexed the age effects with $i = 0, 5, 10, \ldots, 50$ and the periods with $j = 0, 3, 6, \ldots, 45$. The corresponding cohort index is then derived using the standard definition $k = (I - i) + j$, but setting I to the maximum age index of 50. In this work we consider the case in which the interval width of the age groups is M times wider than the interval width of the periods and use the definition of k proposed by Heuer (1997). Suppose age is given in M-year intervals, while period is given on an annual basis. Then $k = M \times (I - i) + j$ and $K = M \times (I - 1) + J$. The grid factor M is generally defined as the ratio of the width of age group and period intervals.

To ensure identifiability of the intercept μ additional constraints must be made. One popular approach is to apply sum-to-zero constraints, also termed the "usual constraints", $\sum_i \theta_i =$

$\sum_j \varphi_j = \sum_k \psi_k = 0$ (Holford, 2005). However, a second identifiability problem remains due to the exact linear relationship of the three time scales which makes it impossible to identify the separate contributions of the individual time effects, i.e. age, period and cohort effects (Holford, 2005).

3 The identifiability problem

The exact linear dependence of age, period and cohort can be viewed as a special case of collinear regressors. Since there are infinitely many linear transformations leading all to the same estimated incidence or mortality rate, it is impossible to estimate a unique set of parameters. (Davison, 2003, p. 398; Yang *et al.*, 2008). In the following we illustrate the identifiability problem for both equally and unequally spaced data, and discuss potential solutions.

3.1 Equally spaced data

Let the interval widths for age and period be equal. Then for an arbitrary $a \in \mathbb{R}$ the linear transformations

$$\theta_i \mapsto \theta_i + a \cdot \left(i - \frac{I+1}{2}\right); \quad \varphi_j \mapsto \varphi_j - a \cdot \left(j - \frac{J+1}{2}\right); \quad \psi_k \mapsto \psi_k + a \cdot \left(k - \frac{K+1}{2}\right) \quad (1)$$

will maintain the sum-to-zero constraints and leave the linear predictor η_{ij} unchanged.

Proof. First, we show that the linear predictor is invariant to transformations of the form (1). Since the linear predictor can be written as

$$\eta_{ij} = \mu + \theta_i + \varphi_j + \psi_k + a \cdot \left(i - \frac{I+1}{2} - j + \frac{J+1}{2} + k - \frac{K+1}{2}\right),$$

we only need to show that the term in brackets is equal to zero:

$$i - \frac{I+1}{2} - j + \frac{J+1}{2} + k - \frac{K+1}{2} = \frac{2i - I - 1 - 2j + J + 1 + 2 \cdot (I - i + j) - I - J)}{2} = 0.$$

Next we show that the sum-to-zero constraints are fulfilled. For the age effects θ_i, $i = 1, \ldots, I$, we obtain:

$$\sum_{i=1}^{I} \left(\theta_i + a \cdot \left(i - \frac{I+1}{2}\right)\right) = \sum_{i=1}^{I} \theta_i + a \sum_{i=1}^{I} i - a\frac{I(I+1)}{2} = 0 + a\frac{I(I+1)}{2} - a\frac{I(I+1)}{2} = 0.$$

The fulfilment of the sum-to-zero constraints for the period effects φ_j, $j = 1, \ldots, J$, and cohort effects ψ_k, $k = 1, \ldots, K$, follows analogously. □

Thus, the rate λ_{ij} can be identified, but age, period and cohort effects cannot. An example to illustrate this invariance is shown in Figure 2. Note that we can rotate the period effects without affecting the model fit, because rotating the period effects in a clockwise direction results in a corresponding rotation of age and cohort effects in a counter-clockwise direction. Thus, the choice of a particular set of parameter estimates is necessarily arbitrary and expert knowledge is required to justify the choice (Holford, 2005). Only non-linear trends, e.g. change points or second differences, are interpretable (Clayton and Schifflers, 1987b). For example in Figure 2

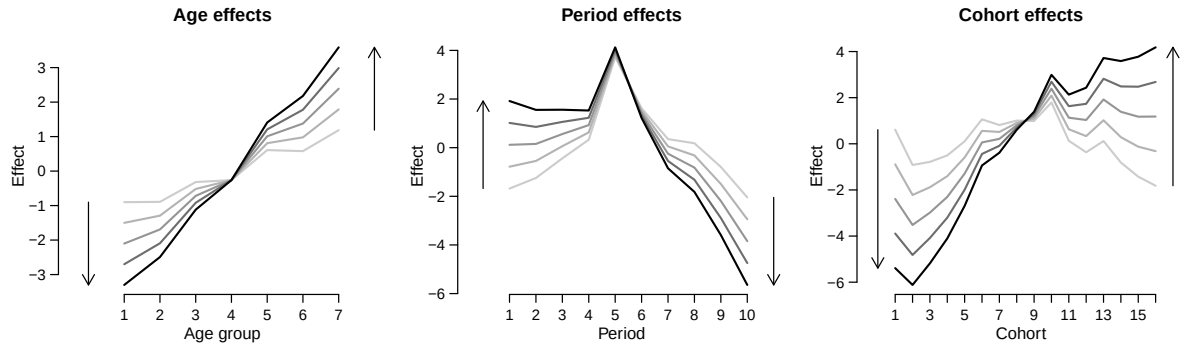

Figure 2: Age, period and cohort effects for different values of a when the interval widths for age and period are equal.

the change points are not affected by the linear transformations, e.g. the change point at period index 5 is always present.

To identify the time effects, i.e. effects for age, period and cohort, a further constraint has to be imposed in addition to the sum-to-zero constraints (Holford, 1983, 1991; Robertson *et al.*, 1999). Over the last decades several proposals have been made to solve the identifiability problem, see for example Fienberg and Mason (1979); Osmond and Gardner (1982); Holford (1983); Robertson and Boyle (1986); Clayton and Schifflers (1987b); Holford (1992); Fu (2000); Kuang *et al.* (2008); Yang *et al.* (2008). One possibility is to limit the attention to two time scales, so that the collinearity present between the three time scales is avoided. However, in applications, two-factor models are frequently not appropriate to fit the data well. Alternatively, one set of parameter effects could be replaced by a more direct measure. For example when analysing lung cancer the period effects could be replaced with a regression variable related to the number of cigarettes sold (Rodgers, 1982; Brown and Kessler, 1988; Knorr-Held and Rainer, 2001). However, to solve the identifiability problem, the covariate should be either included in a parametric fashion or without depending on the time scale for which it was introduced. For example, the identifiability problem remains when using a time-varying non-parametric effect. Suppose we replace the period effects φ_j, $j = 1, \ldots, J$, with smooth time-varying effects β_j for covariate x_j, $j = 1, \ldots, J$:

$$\eta_{ij} = \log \lambda_{ij} = \mu + \theta_i + \beta_j x_j + \psi_k.$$

For $x_j = c$, $c \in \mathbb{R}$, for all $j = 1, \ldots, J$ the standard identifiability problem results. For all other covariate vectors the linear predictor is unchanged for transformations of the form (1), but replacing the transformation of the period effects by:

$$\beta_j x_j \mapsto \beta_j x_j - a\left(j - \frac{J+1}{2}\right) = \underbrace{\left[\beta_j - a\left(\frac{j - (J+1)/2}{x_j}\right)\right]}_{\tilde{\beta}_j} x_j. \tag{2}$$

(The proof is analogous to the case without covariates.)

Provided that not all x_j, $j = 1, \ldots, J$, take the same value, the parameters are "weakly identifiable" within a Bayesian setting because the transformation (2) may change the prior distribution of β_j, which could be for example a random walk of first or second order. By imposing a stochastic constraint the identifiability problem is solved. For instance a random walk of second order will prefer, among all transformations (2), the one where the quadratic second differences are

minimal. In other words, the smooth time-varying effects β_j are kept as linear as possible. Paper IV by Riebler *et al.* (2010b) illustrates the introduction of covariate information on social integration in an APC analysis of Swiss suicide data.

Another approach to solve the identifiability problem is to equate two effects for one of the time scales (Fienberg and Mason, 1979). However, sufficient expert knowledge is required to justify the choice of the parameter pair as the resulting set of parameter estimates is extremely sensitive to it. Osmond and Gardner (1982) impose an additional mathematical constraint to the APC model in which a penalty function is minimised. The function measures the Euclidean distance between the parameters of the two-factor models and those of the full APC model. Then, the solution is given by the parameter estimates that minimise this distance. This approach has been criticised to lack of a theoretical justification (Holford, 1991, 2005). Robertson and Boyle (1986) propose to use more detailed data, namely individual records. They determine the exact date of birth from the exact age and the exact date of death. Subsequently, they divide the compartments of the Lexis diagram into an upper and lower triangular, known as older and younger cohorts respectively, to which the data can be clearly assigned by knowing the date of birth. Since there are now two cohorts associated with each age group i and period j, the exact linear dependence has been solved (Robertson *et al.*, 1999). However, this approach has been controversially discussed as well, see Osmond and Gardner (1989) or Holford (1991).

The adoption of arbitrary constraints can be avoided by inspecting only estimable functions of age, period and cohort effects. Estimable functions are those which do not depend on the constraint imposed to obtain a particular set of parameter estimates. Examples are the linear predictor and, thus, also forecasts, or second differences of time effects (Holford, 2005). Based on the idea of estimable functions, the intrinsic estimator has recently been proposed (Fu, 2000; Yang *et al.*, 2004, 2008). This estimator is a special form of the principal component regression which removes the influence of the null space of the design matrix on the parameter estimates (Yang *et al.*, 2008).

3.2 Unequally spaced data

The identifiability problem described for equally spaced data continues when the age group interval is not equal to the period interval. Transformations of the form

$$\theta_i \mapsto \theta_i + M \times a \cdot \left(i - \frac{I+1}{2}\right); \quad \varphi_j \mapsto \varphi_j - a \cdot \left(j - \frac{J+1}{2}\right); \quad \psi_k \mapsto \psi_k + a \cdot \left(k - \frac{K+1}{2}\right), \tag{3}$$

will maintain the sum-to-zero constraints but the linear predictor is unchanged. Here, M represents the grid factor defined as the ratio of age group and period interval length, see Section 2. The proof is analogous to the case of equally spaced data, but using $K = M \times (I-1) + J$ and $k = M \times (I-i) + j$. Figure 3 shows possible sets of parameter estimates for different values of a that all result in the same linear predictor and fulfil the sum-to-zero constraints.

However, when age and period do not have the same interval width, further identifiability problems arise, see Holford (1983, 2006). For example, let $(J \bmod M) = 0$ and $(K \bmod M) = 0$. For any value of $b_1, \ldots, b_M \in \mathbb{R}$ subject to $\sum_{m=1}^{M} b_m = 0$ the transformations

$$\varphi_j = \varphi_j + b_{1+(j-1) \bmod M} \tag{4}$$

$$\psi_k = \psi_k - b_{1+(k-1) \bmod M} \tag{5}$$

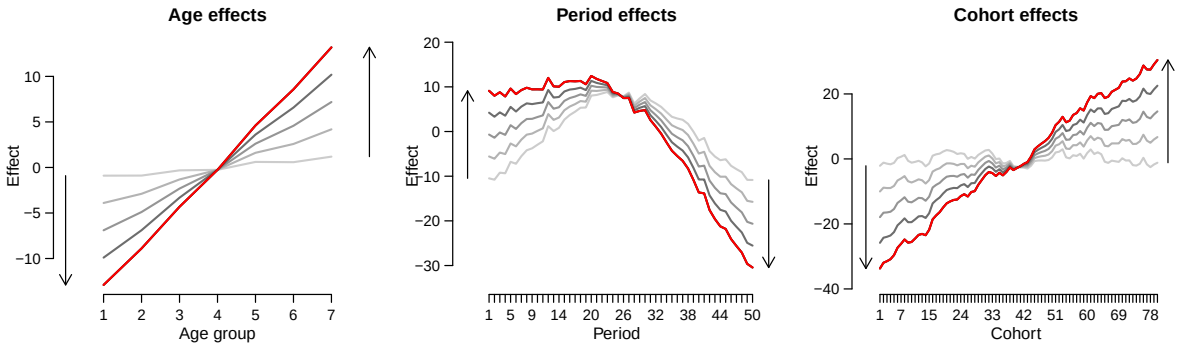

Figure 3: Age, period and cohort effects for different values of a when the interval widths for age and period are not equal.

leave the linear predictor η_{ij} unchanged and will maintain the sum-to-zero constraints for period and cohort effects.

Proof. For $M = 1$ the results follow immediately, since then $b_1 = 0$. Thus let $M > 1$. We first show that the linear predictor η_{ij} is invariant to the transformations (4) and (5).

Since the intercept and the age effects are not affected by the transformations (4) and (5) we only need to show that:

$$b_{1+(j-1) \bmod M} = b_{1+(k-1) \bmod M}.$$

Thus it has to be proven that the indices on both sides are the same. Since $k = M \times (I - i) + j$ it follows that

$$\begin{aligned}
1 + (k-1) \bmod M &= 1 + (M \cdot (I-i) + j - 1) \bmod M \\
&= 1 + (M \cdot (I-i) + j - 1) - \left\lfloor \frac{M \times (I-i) + j - 1}{M} \right\rfloor \cdot M \\
&= M \cdot I - M \cdot i + j - \left\lfloor \frac{M \cdot I - M \cdot i + j - 1}{M} \right\rfloor \cdot M \\
&= j - \left\lfloor \frac{j-1}{M} \right\rfloor \cdot M = 1 + (j-1) \bmod M.
\end{aligned}$$

For the fulfilment of the sum-to-zero constraint, the assumptions $(J \bmod M) = 0$ and $(K \bmod M) = 0$ are essential:

$$\sum_{j=1}^{J} \varphi_j + b_{1+(j-1) \bmod M} = \sum_{j=1}^{J} \varphi_j + \sum_{j=1}^{J} b_{1+(j-1) \bmod M} = \frac{J}{M} \sum_{j=1}^{M} b_{1+(j-1) \bmod M} = \frac{J}{M} \sum_{j=1}^{M} b_j = 0,$$

$$\sum_{k=1}^{K} \psi_k + b_{1+(k-1) \bmod M} = \sum_{k=1}^{K} \psi_k + \sum_{k=1}^{K} b_{1+(k-1) \bmod M} = \frac{K}{M} \sum_{k=1}^{M} b_{1+(k-1) \bmod M} = \frac{K}{M} \sum_{k=1}^{M} b_k = 0.$$

□

Figure 4 shows possible sets of parameter estimates for $M = 5$. The dark-grey curve is obtained from the light-grey curve by transformation (3) using $a = 0.8$. The red curve is derived from the dark-grey curve by applying transformation (4) and (5). Note the five-year periodicity in

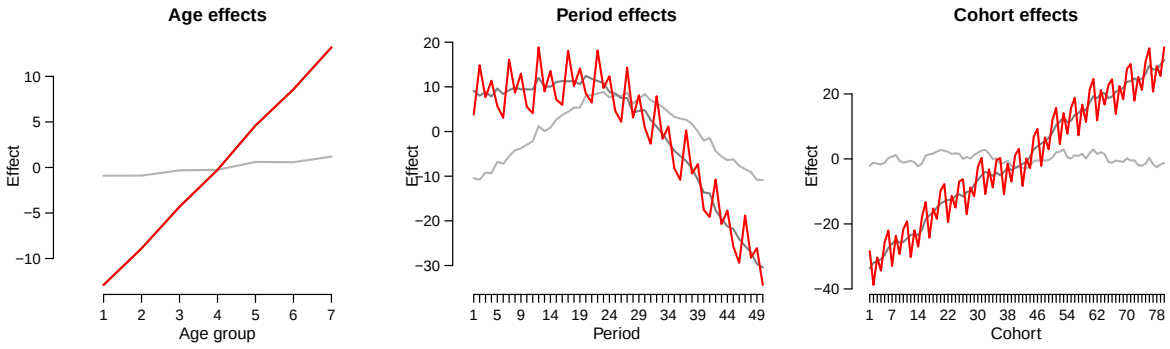

Figure 4: Sets of age, period and cohort effects obtained for an example dataset with $M = 5$ after applying different transformations. All sets result in the same linear predictor and fulfil the sum-to-zero constraints.

the period and cohort estimates of the red curve. The cyclical pattern can lead to extremely misleading conclusions. Cycles that appear in data with unequal intervals must be carefully analysed, especially if they show a periodicity equal to the least common multiple of the interval width of period and cohort (Holford, 2006). Note, however, that Holford (2006) also observed cyclical patterns in an example with $I = 11$, $J = 16$, and where age groups are provided in five-year intervals and period in three-year intervals. Thus, the appearance of cyclical patterns is not specific for datasets with $(J \bmod M) = 0$ and $(K \bmod M) = 0$.

By applying smoothing functions, such as second-order random walks or penalised splines, many of the problems arising for unequally spaced data can be solved. However, the identifiability problem known for equally spaced data remains (Holford, 2006).

4 The multivariate age-period-cohort model

When several registry datasets are available, for example mortality rates for males and females or for different geographical regions, each dataset could be analysed separately by applying univariate APC models. However, depending on the context it might be beneficial to analyse all datasets jointly. Multivariate APC models are used to capture heterogeneous time trends across different strata, for example geographical regions (Hansell *et al.*, 2003; Hansell, 2004; Jacobsen *et al.*, 2004; Riebler and Held, 2010). For comparable strata, it seems plausible that similar factors act on the time scales, so that the time effects corresponding to those time scales may be common. A potentially useful model in many applications would, for example, assume common age effects:

$$y_{ijr}|\eta_{ijr} \sim \text{Po}(n_{ijr}\lambda_{ijr}), \qquad \eta_{ijr} = \log \lambda_{ijr} = \mu_r + \theta_i + \varphi_{jr} + \psi_{kr}. \tag{6}$$

Here, μ_r is the outcome-specific intercept for stratum $r = 1, \ldots, R$, θ_i is the common age effect, and φ_{ir} and ψ_{kr} are outcome-specific period and cohort effects, respectively. As in the univariate case, sum-to-zero constraints are imposed to ensure identifiability of the stratum-specific intercepts: $\sum_i \theta_i = 0$, $\sum_j \varphi_{jr} = \sum_k \psi_{kr} = 0$ for all $r = 1, \ldots, R$. It is straightforward to adapt model formulation (6) to different assumptions, for instance unequal period effects but equal cohort effects. The multivariate APC model inherits all identifiability problems from the univariate APC model. However, differences of stratum-specific time effects are identifiable and

can be interpreted as log relative risks, provided not all time effects vary across strata.

5 Likelihood inference in age-period-cohort models

APC models are usually fitted within a frequentist framework. Age, period and cohort effects are treated as factors and standard Poisson regression is applied (Hansell, 2004; Jacobsen *et al.*, 2004). However, due to the identifiability problems described in Section 3, the columns of the design matrix are linearly dependent, so that the design matrix is not of full rank. Hence, no unique solution exists. At least one additional constraint needs to be imposed (Holford, 1983; Yang *et al.*, 2008). Furthermore, classical maximum likelihood methods will overfit cohorts for which only a single observation exists. For the case in which age group and period intervals are equally spaced, this corresponds to $2 \times R$ cells of the Lexis diagram (Besag *et al.*, 1995). In the unequal case a perfect fit has to be provided to even more cells, namely $2 \times M \times R$ cells. Moreover, problems with zero counts will typically increase since the number of observations on a cohort decreases with increasing grid factor M. A more serious problem noticed in univariate and multivariate APC models is that maximum likelihood estimates tend to be very unstable resulting in artificial cyclical patterns if the widths of age group and period intervals are unequal. However, this problem can be avoided by means of smoothing functions (Holford, 1983, 2006; Riebler and Held, 2010).

In applications, the Poisson assumption may be too restrictive because of greater variability than that of the Poisson distribution. To adjust for such overdispersion, a quasi-likelihood approach is often used (Zheng *et al.*, 1996; Holford, 2006). Here, an overdispersion parameter ϕ equal to the Pearson's χ^2 statistic divided by the residual degrees of freedom is estimated from the regression output to inflate the variances of the regression estimates (Breslow, 1984; McCullagh and Nelder, 1989). An alternative approach replaces the Poisson likelihood with a negative binomial likelihood (Breslow, 1984; Hilbe, 2007). This approach is very similar to the use of random effects which is the usual adjustment for overdispersion within a Bayesian framework. Indeed, if the random effects were log-gamma distributed, they could be integrated out to a negative binomial model leaving only one additional parameter, namely the random effects variance, in the model.

To compare nested models (for example, the age-only model, the age-period model, the age-cohort model and the full APC model) a deviance analysis can be carried out, see for example Clayton and Schifflers (1987a); Thygesen *et al.* (2005). Jacobsen *et al.* (2004) analysed overall mortality of Swedish, Danish and Norwegian women and compared the deviances of three separate univariate APC models with the deviance of the multivariate APC model. Alternatively, or for models that are not nested, the well-known Akaike Information Criterion (AIC) can be computed for model choice (Akaike, 1973). It is defined as

$$\text{AIC} = -2 \log L + 2p,$$

where L is the value of the maximised likelihood function and p is the number of parameters. In applications one computes the AIC value for all candidate models. The model with the lowest AIC value is regarded as the best. Thus, differences in AIC values, rather than absolute AIC values, are important (Burnham and Anderson, 2002). The QAIC criterion represents a modification of AIC derived by principles of quasi-likelihood and is used to adjust for overdispersion (Burnham and Anderson, 2002). QAIC is defined as

$$\text{QAIC} = -2 \log L/\phi + 2p,$$

where ϕ is the estimated overdispersion parameter. If $\phi > 1$, i.e. overdispersion is present, one additional parameter must be added to p.

In Paper II, Held and Riebler (2010) propose a conditional approach for inference in multivariate APC models. This approach results in a multinomial logistic regression model which is estimated based on maximum likelihood inference. Parameters of interest, namely differences of stratum-specific effects, are directly modelled and smoothed, if appropriate. If necessary, quasi-likelihood adjustments for overdispersion based on Pearson's χ^2 statistic for multinomial data are performed (McCullagh and Nelder, 1989, Section 5.5). QAIC is used to compare different multivariate APC models.

6 Bayesian inference in age-period-cohort models

Bayesian APC models incorporate uncertainty about hyperparameters and avoid difficulties arising from the collinearity present in the APC model by the application of mildly informative prior distributions (Besag *et al.*, 1995). In addition, it is straightforward to account for unstructured heterogeneity. Bayesian versions are very popular to project mortality rates, since they do not rely on strong parametric assumptions for future values of period and cohort effects. For example, Osmond (1985) assumed in a frequentist analysis constant age effects and projected period and cohort effects using a linear regression applied to chosen numbers of the most recent period and cohort effects. Bray (2002) provides a discussion of some classical approaches for the projection of mortality or incidence data, and gives a comparison to Bayesian APC models. For applications and discussions on Bayesian APC models we refer to Berzuini and Clayton (1994); Besag *et al.* (1995); Knorr-Held and Rainer (2001); Bray *et al.* (2001); Bray (2002); Baker and Bray (2005); Schmid and Held (2007).

In a Bayesian context, the (multivariate) APC model is a hierarchical model with three stages. The first stage are the data to which a Poisson likelihood is assigned. The second stage is the latent field, here the time effects, i.e. age, period and cohort effects. The third stage are the hyperparameters, typically precisions or correlations.

6.1 Prior assumptions

First, we assign prior distributions to the latent field. Subsequently, prior distributions for the hyperparameters are presented.

Latent Gaussian Markov random field

We start by introducing Gaussian Markov random fields (GMRFs), which will be used throughout this work. A GMRF is a random vector following a multivariate Gaussian distribution with Markov properties. Thus the vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ is called a GMRF with mean $\boldsymbol{\mu}$ and positive definite precision (inverse covariance) matrix $\mathbf{Q}$ if its density has the form:

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2}|\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \mathbf{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

The Markov properties are encoded in the precision matrix $\mathbf{Q}$: $Q_{ij} = 0$, if and only if x_i and x_j are conditionally independent given $\boldsymbol{x}_{-ij} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)^\top$. In most cases only few elements of $\mathbf{Q}$ are nonzero, and $\mathbf{Q}$ is said to be sparse. For further details

see Rue and Held (2005).

The canonical representation of a GMRF $\boldsymbol{x}$ is given by

$$\pi(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x}\right), \tag{7}$$

with canonical parameters $\boldsymbol{b}$ and positive definite precision matrix $\mathbf{Q}$. The mean is $\boldsymbol{\mu} = \mathbf{Q}^{-1}\boldsymbol{b}$. The canonical parametrisation is denoted as $\boldsymbol{x} \sim \mathcal{N}_C(\boldsymbol{b}, \mathbf{Q})$. It is related to the normal distribution via $\mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1}) = \mathcal{N}_C(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$, see Rue and Held (2005, p.27).

In applications, e.g. time series analysis, so called intrinsic GMRFs are of particular interest. Intrinsic GMRFs are always improper, i.e. their precision matrices are not of full rank and they do not have any mean $\boldsymbol{\mu}$. They are used intensively as prior distributions (Rue and Held, 2005, see Chapter 3) and are of the general form

$$\pi(\boldsymbol{x}) = (2\pi)^{-(n-k)/2}(|\mathbf{Q}|^\star)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x}\right).$$

Here, $\mathbf{Q}$ is a symmetric positive semi-definite matrix with rank $n-k$ and $|\mathbf{Q}|^\star$ denotes the generalised determinant of $\mathbf{Q}$, defined as the product of the $n-k$ non-zero eigenvalues. Note that $\boldsymbol{Q}$ no longer represents the precision, since formally the precision does not exist. However, for convenience we will continue to use this terminology. For a general treatment of intrinsic GMRFs we refer to Künsch (1987) and Rue and Held (2005).

In APC models it seems plausible that effects adjacent in time are similar, so that Gaussian priors with mean zero are frequently used for pairwise differences (Besag *et al.*, 1995). For example, for the age effects $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^\top$, a prior based on first-order differences is given by

$$\begin{aligned} f(\boldsymbol{\theta}|\kappa_\theta) &\propto \kappa_\theta^{(I-1)/2} \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=2}^{I}(\theta_i - \theta_{i-1})^2\right) \\ &= \kappa_\theta^{(I-1)/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}^\top \mathbf{R}_\theta^{(1)} \boldsymbol{\theta}\right). \end{aligned}$$

Here, $\mathbf{R}_\theta^{(1)}$ is a precision matrix defined as

$$\mathbf{R}_\theta^{(1)} = \kappa_\theta \begin{pmatrix} 1 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & -1 & 2 & -1 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & -1 & 2 & -1 & & \\ & & & & -1 & 2 & -1 & \\ & & & & & -1 & 2 & -1 \\ & & & & & & -1 & 1 \end{pmatrix}$$

where κ_θ is the precision parameter that determines the degree of smoothing. The higher the precision, the smoother the estimated parameter vector. Note that non-given entries in $\mathbf{R}_\theta^{(1)}$ are zero. This prior corresponds to the directed formulation as a random walk of first order (RW1) $\theta_i|\theta_{i-1}, \ldots, \theta_1 \sim \mathcal{N}(\theta_{i-1}, \kappa_\theta^{-1})$, $i = 2, \ldots, I$ using a flat prior for θ_1.

Similarly, a prior based on second-order differences is given by

$$f(\boldsymbol{\theta}|\kappa_\theta) \propto \kappa_\theta^{(I-2)/2} \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=3}^{I}((\theta_i-\theta_{i-1})-(\theta_{i-1}-\theta_{i-2}))^2\right)$$
$$= \kappa_\theta^{(I-2)/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}^\top \mathbf{R}_\theta^{(2)} \boldsymbol{\theta}\right)$$

with

$$\mathbf{R}_\theta^{(2)} = \kappa_\theta \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & 1 & -4 & 5 & -2 \\ & & & & & 1 & -2 & 1 \end{pmatrix}. \tag{8}$$

The equivalent directed formulation is a random walk of second order (RW2) $\theta_i|\theta_{i-1},\ldots,\theta_1 \sim \mathcal{N}(2\theta_{i-1}-\theta_{i-2},\kappa_\theta^{-1})$, $i=3,\ldots,I$ using independent uniform priors both for θ_1 and θ_2. While RW1 penalises deviations from a constant, RW2 penalises deviations from a linear trend $\theta_i = 2\theta_{i-1}-\theta_{i-2}$, $i=3,\ldots,I$. Random walks are closely related to smoothing splines. For instance, the RW2 represents a discrete-time analogue of a cubic smoothing spline (Fahrmeir and Tutz, 2001). Since the RW2 penalises the second differences which are identifiable in APC models this prior is particularly attractive in our context. Both RW1 and RW2 are prominent examples of intrinsic GMRFs.

In the context of structured additive regression, smoothing priors may be applied to all time effects of the (multivariate) APC model and the estimated latent parameters may be analysed directly (Fahrmeir and Lang, 2001; Fahrmeir and Tutz, 2001). However, due to the implicit identifiability problem present in APC models, this procedure is not applicable unless additional constraints are imposed. The use of a RW1 for all time effects would solve the identifiability problem. However, this is only due to a constraint, not deterministic but stochastic. Most importantly, if the main objects of interest are identifiable, latent parameters need not be identifiable in a Bayesian setting (Besag *et al.*, 1995, rejoinder). In the case of multivariate APC models, the parameters of interest are the identifiable differences of stratum-specific effects.

Applying separate random walk priors to stratum-specific time effects cannot capture a potential dependence between them. Since the time scales might be subject to the same environmental influence it seems plausible to link them. In Paper III, Riebler *et al.* (2010a) propose the use of correlated GMRF priors instead of independent smoothing priors. In multiple time series analysis multivariate random walk priors are frequently used (Harvey, 1990). However, up to now they have not been applied in the context of multivariate APC models. Riebler *et al.* (2010a) propose a Kronecker product precision matrix $\mathbf{C}^{-1}\otimes\mathbf{R}^{(2)}$ where $\mathbf{C}$ is a uniform correlation matrix with 1s on the diagonal and unknown correlation parameter ρ on all remaining entries, and $\mathbf{R}^{(2)}$ is the precision matrix of the univariate RW2 given in (8). This formulation corresponds to a multivariate RW2 with correlated increments.

To account for additional "unstructured" heterogeneity which can be explained by neither age, period nor cohort effects, we include additional outcome-specific random parameters z_{ijr} into the linear predictor (6). Typically these parameters are assumed to be independent and normally distributed with mean zero and unknown variance κ_z^{-1} (Besag *et al.*, 1995). However, to account

for extra-Poisson variation induced by correlated unobserved factors a correlated formulation across all strata seems plausible, so that $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijR})^\top \sim \mathcal{N}(\mathbf{0}, \kappa_z^{-1}\mathbf{C}_z)$, where $\mathbf{C}_z$ represents a uniform correlation matrix.

The introduction of the additional overdispersion parameters z_{ijr} has the advantage that the model can be reparameterised following Besag *et al.* (1995). Thus, the full conditional distributions of all latent parameters are standard distributions and Gibbs sampling can be applied.

For the intercepts independent flat priors $p(\mu_r) \propto$ const. will be used.

Hyperparameters

In a fully Bayesian analysis, all hyperparameters are treated as unknown and will be estimated simultaneously in the model. Thus hyperpriors for the precisions and for the correlation parameters (if present) need to be assigned. Note, that the precisions of the time effects can be stratum-dependent or stratum-independent. In Paper I, we chose stratum-dependent precision parameters, while in Paper III and IV stratum-independent precision parameters were used.

We assign to all precision parameters weakly informative but proper gamma distributions $\mathrm{Ga}(a, b)$ to avoid problems with improper hyperpriors. Problems caused by improper hyperpriors are for example discussed in Hobert and Casella (1996). The gamma distribution $\mathrm{Ga}(a, b)$ is defined as in Bernardo and Smith (1994) with density function:

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \quad \text{with } x > 0 \text{ and } a, b > 0.$$

Gamma hyperpriors for the precisions are computationally convenient since the implied full conditional distribution is again a gamma distribution. As suggested by Knorr-Held and Rainer (2001) we use $a = 1, b = 0.000\,05$ for the precisions of the time effects and $a = 1, b = 0.005$ for precision κ_z of the overdispersion, as we expect a slightly larger variation in the overdispersion parameters.

Correlation parameters ρ are reparameterised using the general Fisher's z-transformation (Fisher, 1958, page 219):

$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} \qquad \rho^\star = \log\left(\frac{1 + \rho \cdot (R-1)}{1 - \rho}\right), \tag{9}$$

where $\rho^\star$ can take values over the whole real line. We assign a normal prior with mean zero and fixed precision $\kappa_{\rho^\star}$ to $\rho^\star$. Note that the general Fisher's z-transformation (9) ensures that ρ only takes values within $(-1/(R-1), 1)$, since

$$\lim_{\rho^\star \to -\infty} \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} = -\frac{1}{R-1},$$
$$\lim_{\rho^\star \to \infty} \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} = \lim_{\rho^\star \to \infty} \frac{1 - \frac{1}{\exp(\rho^\star)}}{1 + \frac{R}{\exp(\rho^\star)} - \frac{1}{\exp(\rho^\star)}} = 1.$$

Thus, the correlation matrix $\mathbf{C}$ is ensured to be positive definite. The prior probability that ρ is larger than 0 is 0.5 independent of R, which is convenient for testing a posteriori whether ρ is larger than 0.

6.2 Markov chain Monte Carlo

Within a Bayesian hierarchical framework, the standard method for parameter estimation is Markov chain Monte Carlo (MCMC). When accounting for unstructured heterogeneity, the (multivariate) APC model can be reparameterised as proposed by Besag *et al.* (1995). Then the full conditional distributions for all parameters (except the correlations) are standard distributions, so that Gibbs sampling can be applied, see Appendix II. Each set of time effects $\boldsymbol{\theta}_{(r)} = (\theta_{1(r)}, \ldots, \theta_{I(r)})^\top$, $\boldsymbol{\varphi}_{(r)} = (\varphi_{1(r)}, \ldots, \varphi_{J(r)})^\top$ and $\boldsymbol{\psi}_{(r)} = (\psi_{1(r)}, \ldots, \psi_{K(r)})^\top$, $r = 1, \ldots, R$, is updated as a block with correct incorporation of the sum-to-zero constraint instead of so called "centring on-the-fly".

The full conditional of the linear predictor is a non-standard distribution. To sample from this full conditional distribution we consider two approaches: 1) A Metropolis-Hastings algorithm using an appropriate proposal distribution based on a GMRF approximation (Rue and Held, 2005, Section 4.4), and 2) improved auxiliary mixture sampling (Frühwirth-Schnatter *et al.*, 2009).

GMRF approximation

Uncorrelated overdispersion parameters

In the case of multivariate APC models, e.g. (6), with uncorrelated overdispersion parameters we approximate the log full conditional of the linear predictor using a second-order Taylor approximation. Given

$$y_{ijr} \sim \text{Po}(n_{ijr}\lambda_{ijr}), \qquad \xi_{ijr} = \log(\lambda_{ijr}) = \underbrace{\mu_r + \theta_i + \varphi_{jr} + \psi_{kr}}_{\eta_{ijr}} + z_{ijr},$$

it follows that $\xi_{ijr} \sim \mathcal{N}(\eta_{ijr}, \kappa_z^{-1})$, so that the full conditional of the linear predictor is

$$\pi(\xi_{ijr}|\cdot) \propto \exp\left(-\frac{\kappa_z}{2}(\xi_{ijr} - \eta_{ijr})^2 + y_{ijr}(\log(n_{ijr}) + \xi_{ijr}) - n_{ijr}\exp(\xi_{ijr})\right) = \exp(f(\xi_{ijr})). \quad (10)$$

In order to approximate $f(\xi_{ijr})$ it is common to use the second-order Taylor expansion of the (unnormalised) log full conditional distribution around a suitable value $\xi_{ijr}^{(0)}$,

$$\begin{aligned} f(\xi_{ijr}) &\approx f(\xi_{ijr}^{(0)}) + f'(\xi_{ijr}^{(0)})(\xi_{ijr} - \xi_{ijr}^{(0)}) + \frac{1}{2}f''(\xi_{ijr}^{(0)})(\xi_{ijr} - \xi_{ijr}^{(0)})^2 \\ &= a + b\,\xi_{ijr} - \frac{1}{2}c\,\xi_{ijr}^2, \end{aligned} \quad (11)$$

where $b = f'(\xi_{ijr}^{(0)}) - f''(\xi_{ijr}^{(0)})\xi_{ijr}^{(0)}$ and $c = -f''(\xi_{ijr}^{(0)})$. The value of a is not relevant as it does not depend on ξ_{ijr}, but only on $\xi_{ijr}^{(0)}$. The full conditional distribution can now be approximated by

$$\tilde{\pi}(\xi_{ijr}|\cdot) \propto \exp\left(-\frac{1}{2}c\,\xi_{ijr}^2 + b\,\xi_{ijr}\right), \quad (12)$$

where $\tilde{\pi}(\cdot|\cdot)$ is an approximated (conditional) density of its arguments. Equation (12) corresponds to the canonical representation $\mathcal{N}_C(b, c)$, see (7), of a normal distribution with mean b/c and variance c^{-1}. Figure 5 shows the Taylor approximation for $y_{ijr} = 3$, $\eta_{ijr} = 0$ and $\kappa_z = 0.01$.

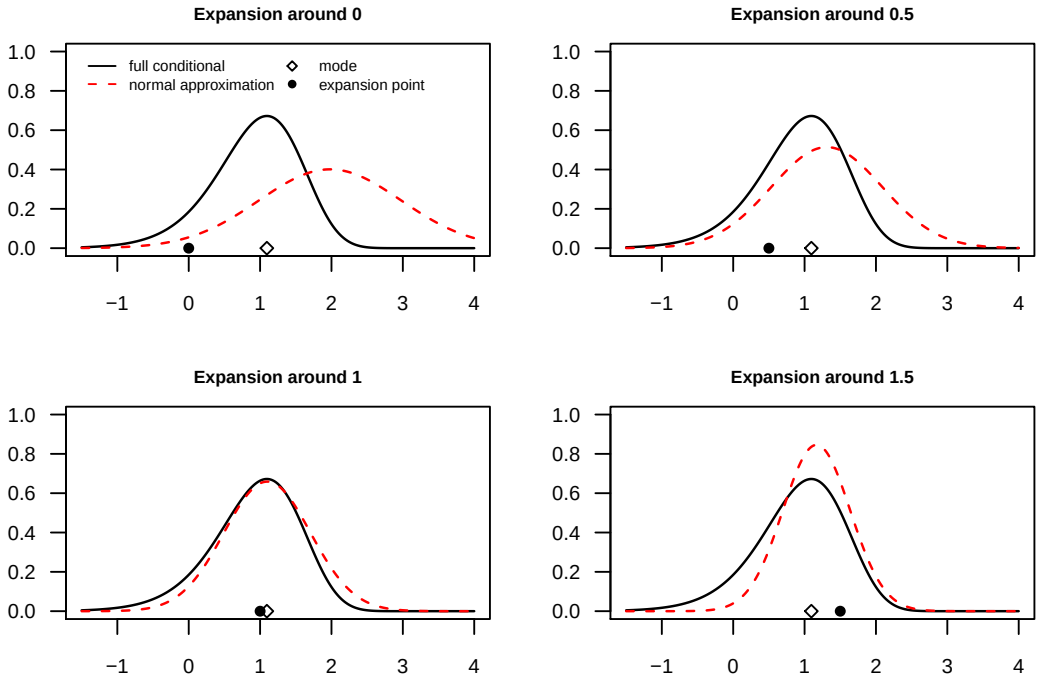

Figure 5: Normal approximation (dashed line) of the full conditional distribution (10) of the linear predictor (solid line) for $y_{ijr} = 3$, $\eta_{ijr} = 0$ and $\kappa_z = 0.01$ based on a second-order Taylor expansion around $\xi_{ijr}^{(0)} = 0, 0.5, 1.0$ and 1.5 (upper left to lower right). The value of $\xi_{ijr}^{(0)}$ is indicated by a $\bullet$ and the mode of the full conditional by a $\diamond$ in each plot.

As seen in the plots, the approximation is better, the closer the expansion point $\xi_{ijr}^{(0)}$ is to the mode of the full conditional distribution $\pi(\xi_{ijr}|\cdot)$ (Rue and Held, 2005). Now, this normal distribution can be used as a proposal distribution $\tilde{\pi}(\xi_{ijr}^{\star}|\xi_{ijr}^{(0)})$ in the Metropolis-Hastings algorithm (Gilks *et al.*, 1996). For $\xi_{ijr}^{(0)}$ the current value of the Markov chain can be used. The acceptance probability of a new value $\xi_{ijr}^{\star}$ is

$$\alpha = \min \left\{ 1, \frac{\pi(\xi_{ijr}^{\star}|\cdot)}{\pi(\xi_{ijr}^{(0)}|\cdot)} \frac{\tilde{\pi}(\xi_{ijr}^{(0)}|\xi_{ijr}^{\star})}{\tilde{\pi}(\xi_{ijr}^{\star}|\xi_{ijr}^{(0)})} \right\}.$$

Note, that for the calculation of the acceptance probability, not only the computation of $\tilde{\pi}(\xi_{ijr}^{\star}|\xi_{ijr}^{(0)})$, but also the computation of $\tilde{\pi}(\xi_{ijr}^{(0)}|\xi_{ijr}^{\star})$ is needed. Hence a second Taylor approximation around the expansion point $\xi_{ijr}^{\star}$ has to be computed and evaluated at the value $\xi_{ijr}^{(0)}$. In total there are $2 \times R \times I \times J$ second-order Taylor expansions needed per MCMC iteration.

Correlated overdispersion parameters

In the case of multivariate APC models with correlated overdispersion parameters the full conditional of the linear predictor $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \ldots, \xi_{ijR})^\top$ is given by:

$$\pi(\boldsymbol{\xi}_{ij}|\cdot) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})^\top \{\kappa_z \mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij}) + \sum_{r=1}^{R} \underbrace{[y_{ijr}(\log(n_{ijr}) + \xi_{ijr}) - n_{ijr}\exp(\xi_{ijr})]}_{f(\xi_{ijr})}\right)$$

where $\boldsymbol{\eta}_{ij} = (\eta_{ij1}, \ldots, \eta_{ijR})^\top$. In this context, it is useful to approximate the log likelihood using a second-order Taylor expansion and to use this GMRF approximation as a Metropolis-Hastings proposal in the MCMC algorithm. We approximate each $f(\xi_{ijr})$, $r = 1, \ldots, R$, using a quadratic Taylor expansion (11) and use these approximations to construct a suitable GMRF proposal density $\tilde{\pi}(\boldsymbol{\xi}_{ij}|\cdot)$ for $\pi(\boldsymbol{\xi}_{ij}|\cdot)$:

$$\begin{aligned}\tilde{\pi}(\boldsymbol{\xi}_{ij}|\cdot) &\propto \exp\left(-\frac{1}{2}\boldsymbol{\xi}_{ij}^\top \{\kappa_z \mathbf{C}_z^{-1}\}\boldsymbol{\xi}_{ij} + \boldsymbol{\eta}_{ij}^\top \{\kappa_z \mathbf{C}_z^{-1}\}\boldsymbol{\xi}_{ij} + \sum_{r=1}^{R}(a_r + b_r\,\xi_{ijr} - \frac{1}{2}c_r\,\xi_{ijr}^2)\right) \\ &\propto \exp\left(-\frac{1}{2}\boldsymbol{\xi}_{ij}^\top \{\kappa_z \mathbf{C}_z^{-1} + \mathrm{diag}(\boldsymbol{c})\}\boldsymbol{\xi}_{ij} + (\kappa_z \mathbf{C}_z^{-1}\boldsymbol{\eta}_{ij} + \boldsymbol{b})^\top \boldsymbol{\xi}_{ij}\right).\end{aligned}$$

The canonical representation is

$$\mathcal{N}_C(\kappa_z \mathbf{C}_z^{-1}\boldsymbol{\eta}_{ij} + \boldsymbol{b}, \kappa_z \mathbf{C}_z^{-1} + \mathrm{diag}(\boldsymbol{c})).$$

This GMRF can be used, analogously to the uncorrelated case, as a proposal distribution in the Metropolis-Hastings update of $\boldsymbol{\xi}_{ij}$.

The sampling scheme:

1. For $\xi_{ijr}^{(0)}$, $r = 1, \ldots, R$ take the values $\xi_{ij1}, \ldots, \xi_{ijR}$ of the current vector $\boldsymbol{\xi}_{ij}$ of the simulated Markov chain. Calculate a second-order Taylor approximation around each value $\xi_{ij1}^{(0)}, \ldots, \xi_{ijR}^{(0)}$.
2. Propose a new vector $\boldsymbol{\xi}_{ij}^\star$ drawn from the proposal distribution:

$$\mathcal{N}_C(\kappa_z \mathbf{C}_z^{-1}\boldsymbol{\eta}_{ij} + \boldsymbol{b}_0, \kappa_z \mathbf{C}_z^{-1} + \mathrm{diag}(\boldsymbol{c}_0)) \tag{13}$$

3. Determine the density of (13) at the vector $\boldsymbol{\xi}_{ij}^\star$ to evaluate $\tilde{\pi}(\boldsymbol{\xi}_{ij}^\star|\boldsymbol{\xi}_{ij}^{(0)})$.
4. Construct a univariate Taylor expansion around each value $\xi_{ij1}^\star, \ldots, \xi_{ijR}^\star$.
5. Determine the density of the resulting multivariate normal distribution at the current vector $\boldsymbol{\xi}_{ij}^{(0)}$ to evaluate $\tilde{\pi}(\boldsymbol{\xi}_{ij}^{(0)}|\boldsymbol{\xi}_{ij}^\star)$.
6. Accept the proposal with:

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{\xi}_{ij}^\star|\cdot)}{\pi(\boldsymbol{\xi}_{ij}^{(0)}|\cdot)} \frac{\tilde{\pi}(\boldsymbol{\xi}_{ij}^{(0)}|\boldsymbol{\xi}_{ij}^\star)}{\tilde{\pi}(\boldsymbol{\xi}_{ij}^\star|\boldsymbol{\xi}_{ij}^{(0)})}\right\}.$$

This procedure has to be performed for each vector $\boldsymbol{\xi}_{ij}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$. Hence,

$2 \times R \times I \times J$ second-order Taylor expansions need to be calculated in total. Per iteration $2 \times I \times J$ different R-variate GMRFs are used.

In our applications the acceptance rates were very high. Hence, it was not necessary to apply the Taylor approximation iteratively or to use line search algorithms in order to locate the mode of the log likelihood. For more details on these methods, we refer to Gamerman (1997) and Rue and Held (2005, Section 4.4.1).

Auxiliary mixture sampling

Alternatively to the GMRF approximation, in Paper I Riebler and Held (2010) applied the improved auxiliary mixture sampling approach proposed by Frühwirth-Schnatter *et al.* (2009) to estimate multivariate (uncorrelated) APC models. This estimation method assures that all full conditional distributions including $f(\xi_{ijr}|\cdot)$ are standard distributions, so that Gibbs sampling can be applied (Gilks *et al.*, 1996). In the case of Poisson data the approach induces at most two auxiliary variables for each observation. The expectation of these latent parameters is a linear function of the unknown parameters and their error distributions are approximated through a finite number of Gaussian mixture components.

Consider again (6), say, so that

$$y_{ijr} \sim \text{Po}(\underbrace{n_{ijr}\lambda_{ijr}}_{\lambda'_{ijr}}), \qquad \xi'_{ijr} = \log(\lambda'_{ijr}) = \underbrace{\log(n_{ijr}) + \mu_r + \theta_i + \varphi_{jr} + \psi_{kr}}_{\eta'_{ijr}} + z_{ijr}.$$

Note that we included the number of persons at risk n_{ijr} as offset $\log(n_{ijr})$ into the linear predictor.

Following Frühwirth-Schnatter *et al.* (2009) the distribution of $y_{ijr}|\lambda'_{ijr}$ can be regarded as the distribution of the number of jumps of an unobserved Poisson process with intensity λ'_{ijr}, having occurred in the unit time interval $(0, 1)$ (Grimmett and Stirzaker, 2001, Section 6.8). The first step of data augmentation introduces for each observation $y_{ijr} > 0$ the arrival time of the last jump before 1 denoted by $t^\star_{ijr,2}$ and the inter-arrival time between the last jump before and the first jump after 1 denoted by $t^\star_{ijr,1}$. Figure 6 illustrates the Poisson process.

It is known that $t^\star_{ijr,2}$ follows a $\text{Ga}(y_{ijr}, \lambda'_{ijr})$ distribution

$$t^\star_{ijr,2} = \frac{\delta_{ijr,2}}{\lambda'_{ijr}}, \qquad \delta_{ijr,2} \sim \text{Ga}(y_{ijr}, 1) \tag{14}$$

and $t^\star_{ijr,1}$ an exponential distribution

$$t^\star_{ijr,1} = \frac{\delta_{ijr,1}}{\lambda'_{ijr}}, \qquad \delta_{ijr,1} \sim \text{Exp}(1). \tag{15}$$

We can reformulate equations (15) and (14) into

$$-\log t^\star_{ijr,1} = \xi'_{ijr} + \epsilon_{ijr,1} \tag{16}$$

$$-\log t^\star_{ijr,2} = \xi'_{ijr} + \epsilon_{ijr,2} \tag{17}$$

where $\epsilon_{ijr,1} = -\log \delta_{ijr,1}$ with $\delta_{ijr,1} \sim \text{Exp}(1) = \text{Ga}(1, 1)$ and $\epsilon_{ijr,2} = -\log \delta_{ijr,2}$ with $\delta_{ijr,2} \sim \text{Ga}(y_{ij}, 1)$. For nonzero observations we introduce the bivariate latent variable $t_{ijr} = (t^\star_{ijr,1}, t^\star_{ijr,2})$ while for zero observations only the single latent variable $t_{ijr} = t^\star_{ijr,1}$ is needed.

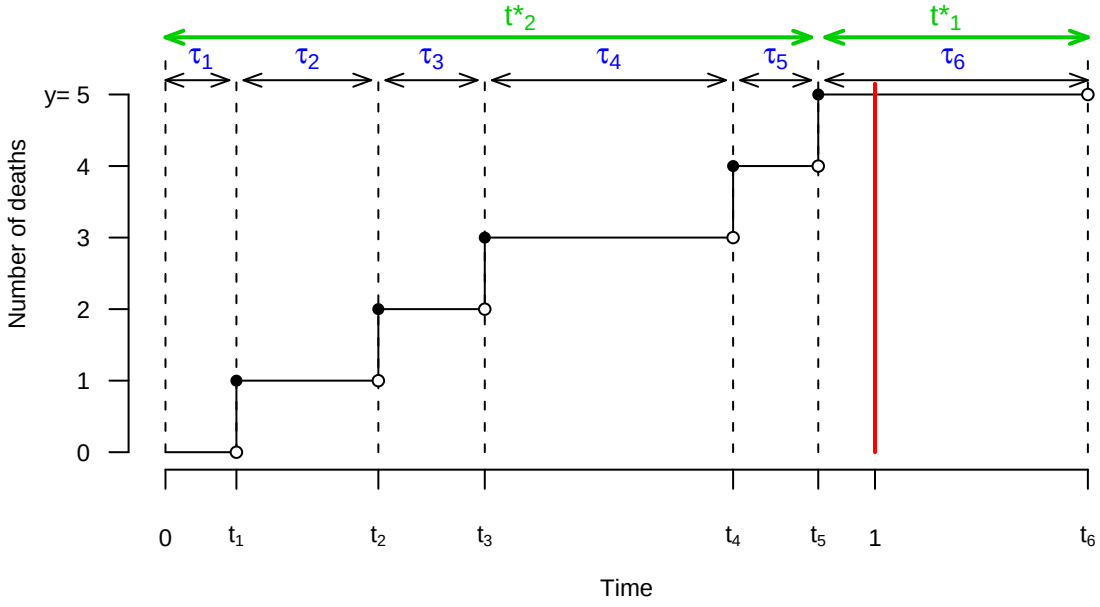

Figure 6: Poisson process. The arrival times are denoted by $t_1, \ldots, t_6$ and the inter-arrival times by $\tau_1, \ldots, \tau_6$. The two auxiliary mixture variables are $t_2^\star$ and $t_1^\star$.

In the second step of data augmentation, the distributions of $\epsilon_{ijr,1}$ and $\epsilon_{ijr,2}$ in (16) and (17) are approximated by a (finite) mixture of normal distributions.

The error term $\epsilon_{ijr,1}$ in equation (16) follows the negative of a log Exp(1) distribution, whose density is independent of any unknown model parameters and can be written as

$$p(\epsilon_{ijr,1}) = \exp(-\epsilon_{ijr,1} - \exp(-\epsilon_{ijr,1})), \quad \epsilon_{ijr,1} \in \mathbb{R}.$$

To obtain a model that is conditionally Gaussian, this non-normal density is approximated by a mixture of U normal components with mean m_u and variance s_u^2 for the u-th component with $u = 1, \ldots, U$:

$$p(\epsilon_{ijr,1}) = \exp(-\epsilon_{ijr,1} - \exp(-\epsilon_{ijr,1})) \approx \sum_{u=1}^{U} w_u \mathcal{N}(\epsilon_{ijr,1}; m_u, s_u^2), \tag{18}$$

where $\mathcal{N}(\epsilon_{ijr,1}; m_u, s_u^2)$ denotes the Gaussian density. Each normal component u is weighted by a specific weight w_u. By minimising the Kullback-Leibler distance between the true density distribution and the mixture approximation, the appropriate parameters (w_u, m_u, s_u^2) were determined numerically for $U = 2, \ldots, 10$, see Frühwirth-Schnatter and Frühwirth (2007).

The error term $\epsilon_{ijr,2}$ in (17) follows a negative log gamma distribution with integer shape parameter ν equal to y_{ijr}. A Gaussian mixture approximation for arbitrary ν is given by

$$p(\epsilon_{ijr,2}, \nu) = \frac{\exp(-\nu\epsilon_{ijr,2} - \exp(\epsilon_{ijr,2}))}{\Gamma(\nu)} \approx \sum_{u=1}^{U(\nu)} w_u(\nu) \mathcal{N}(\epsilon_{ijr,2}; m_u(\nu), s_u^2(\nu)),$$

where the number of necessary components $U(\nu)$, the weights $w_u(\nu)$, the mean values $m_u(\nu)$

and the variances $s_u^2(\nu)$ depend on ν. For $\nu = 1$ the same mixture approximation as in equation (18) is obtained.

By introducing and conditioning on the bivariate latent variable t_{ijr}, and the latent component indicators $u_{ijr} = (u_{ijr,1}, u_{ijr,2})$ for each observation y_{ijr} the model reduces to a linear Gaussian model

$$\begin{aligned} -\log t^\star_{ijr,1} &= \xi'_{ijr} + m_{u_{ijr,1}}(1) + \epsilon_{ijr,1}, & \epsilon_{ijr,1}|u_{ijr,1} &\sim \mathcal{N}(0, s^2_{u_{ijr,1}}(1)), \\ -\log t^\star_{ijr,2} &= \xi'_{ijr} + m_{u_{ijr,2}}(y_{ijr}) + \epsilon_{ijr,2}, & \epsilon_{ijr,2}|u_{ijr,2} &\sim \mathcal{N}(0, s^2_{u_{ijr,2}}(y_{ijr})). \end{aligned}$$

For $y_{ijr} = 0$ only the first equation is necessary.

The full conditional distribution for the linear predictor ξ'_{ijr} can now be expressed as

$$\begin{aligned} f(\xi'_{ijr}|\cdot) &\propto f(\xi'_{ijr}|\eta'_{ijr}, \kappa_z) f(-\log t^\star_{ijr,1}|\xi'_{ijr}, m_{u_{ijr,1}}(1), s^2_{u_{ijr,1}}(1)) \\ &\quad \cdot f(-\log t^\star_{ijr,2}|\xi'_{ijr}, m_{u_{ijr,2}}(y_{ijr}), s^2_{u_{ijr,2}}(y_{ijr}))^{\mathbb{1}(y_{ijr}>0)} \\ &\propto \exp\Bigg(-\frac{\kappa_z}{2}(\xi'_{ijr} - \eta'_{ijr})^2 - \frac{(s^2_{u_{ijr,1}}(1))^{-1}}{2}\big[-\log t^\star_{ijr,1} - (\xi'_{ijr} + m_{u_{ijr,1}}(1))\big]^2 \\ &\quad - \Big\{\frac{(s^2_{u_{ijr,2}}(y_{ijr}))^{-1}}{2}\big[-\log t^\star_{ijr,2} - (\xi'_{ijr} + m_{u_{ijr,2}}(y_{ijr}))\big]^2\Big\} \cdot \mathbb{1}(y_{ijr} > 0)\Bigg) \\ &= \exp\Bigg(-\frac{\kappa_z}{2}(\xi'_{ijr} - \eta'_{ijr})^2 - \frac{(s^2_{u_{ijr,1}}(1))^{-1}}{2}\big[\xi'_{ijr} - (-\log t^\star_{ijr,1} - m_{u_{ijr,1}}(1))\big]^2 \\ &\quad - \Big\{\frac{(s^2_{u_{ijr,2}}(y_{ijr}))^{-1}}{2}\big[\xi'_{ijr} - (-\log t^\star_{ijr,2} - m_{u_{ijr,2}}(y_{ijr}))\big]^2\Big\} \cdot \mathbb{1}(y_{ijr} > 0)\Bigg) \end{aligned}$$

so that $\xi'_{ijr}|\cdot \sim \mathcal{N}(C_\xi^{-1} c_\xi, C_\xi^{-1})$ where

$$\begin{aligned} C_\xi &= \kappa_z + (s^2_{u_{ijr,1}}(1))^{-1} + (s^2_{u_{ijr,2}}(y_{ijr}))^{-1} \cdot \mathbb{1}(y_{ijr} > 0) \\ c_\xi &= \kappa_z \eta'_{ijr} + (s^2_{u_{ijr,1}}(1))^{-1}(-\log t^\star_{ijr,1} - m_{u_{ijr,1}}(1)) \\ &\qquad + (s^2_{u_{ijr,2}}(y_{ijr}))^{-1}(-\log t^\star_{ijr,2} - m_{u_{ijr,2}}(y_{ijr})) \cdot \mathbb{1}(y_{ijr} > 0) \end{aligned}$$

with $\mathbb{1}(y_{ijr} > 0)$ specifying the indicator function, which takes the value 1 for $y_{ijr} > 0$ and the value 0 for $y_{ijr} = 0$.

The sampling scheme:

Let us consider model (6), say. Select starting values for all precisions, for $\boldsymbol{t} = (t_{111}, \ldots, t_{I11}, t_{121}, \ldots, t_{I21}, \ldots, t_{IJ1}, , t_{112}, \ldots, t_{IJR})^\top$ and $\boldsymbol{u} = (u_{111}, \ldots, u_{I11}, u_{121}, \ldots, u_{I21}, \ldots, u_{IJ1}, , u_{112}, \ldots, u_{IJR})^\top$ and for the main effects $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^\top$, $\tilde{\boldsymbol{\varphi}} = (\varphi_{11}, \ldots, \varphi_{J1}, \varphi_{12}, \ldots, \varphi_{J2}, \ldots, \varphi_{JR})^\top$, $\tilde{\boldsymbol{\psi}} = (\psi_{11}, \ldots, \psi_{K1}, \psi_{12}, \ldots, \psi_{K2}, \ldots, \psi_{KR})^\top$.

Repeat the following steps:

1. Sample $\boldsymbol{\xi}' = (\xi'_{111}, \ldots, \xi'_{I11}, \xi'_{121}, \ldots, \xi'_{I21}, \ldots, \xi'_{IJ1}, \xi'_{112}, \ldots, \xi'_{IJR})^\top$ conditional on $\boldsymbol{t}, \boldsymbol{u}, \boldsymbol{\mu} = (\mu_1, \ldots, \mu_r)^\top, \boldsymbol{\theta}, \tilde{\boldsymbol{\varphi}}, \tilde{\boldsymbol{\psi}}, \kappa_z$ and $\boldsymbol{y} = (y_{111}, \ldots, y_{I11}, y_{121}, \ldots, y_{I21}, \ldots, y_{IJ1}, y_{112}, \ldots, y_{IJR})^\top$.

2. Sample the inter-arrival times $\boldsymbol{t}$ and the component indicators $\boldsymbol{u}$ conditional on $\boldsymbol{\xi}', \boldsymbol{\mu}, \boldsymbol{\theta}, \tilde{\boldsymbol{\varphi}}, \tilde{\boldsymbol{\psi}}$ and $\boldsymbol{y}$ by running the following steps for $i = 1, \ldots, I$, $j = 1, \ldots, J$, $r = 1, \ldots, R$:
   a) Sample $\delta_{ijr} \sim \text{Exp}(\lambda'_{ijr})$. If $y_{ijr} = 0$, set $t^{\star}_{ijr,1} = 1 + \delta_{ijr}$. If $y_{ijr} > 0$, sample $t^{\star}_{ijr,2}$ from a $\text{Be}(y_{ijr}, 1)$-distribution and set $t^{\star}_{ijr,1} = 1 - t^{\star}_{ijr,2} + \delta_{ijr}$.
   b) Sample the component indicator $u_{ijr,1}$ from the following discrete distribution where $g = 1, \ldots, U(1)$:

$$P(u_{ijr,1} = g | t^{\star}_{ijr,1}, \xi'_{ijr}) \propto \frac{w_g(1)}{s_g(1)} \exp\left\{ -\frac{1}{2} \left(\frac{-\log t^{\star}_{ijr,1} - \xi'_{ijr} - m_g(1)}{s_g(1)} \right)^2 \right\}.$$

   If $y_{ijr} > 0$, sample the component indicators $u_{ijr,2}$ from the following discrete distribution where $g = 1, \ldots, U(y_{ijr})$:

$$P(u_{ijr,2} = g | t^{\star}_{ijr,2}, \xi'_{ijr}, y_{ijr}) \propto \frac{w_g(y_{ijr})}{s_g(y_{ijr})} \exp\left\{ -\frac{1}{2} \left(\frac{-\log t^{\star}_{ijr,2} - \xi'_{ijr} - m_g(y_{ijr})}{s_g(y_{ijr})} \right)^2 \right\}.$$

   c) Update $\boldsymbol{\mu}, \boldsymbol{\theta}, \tilde{\boldsymbol{\varphi}}, \tilde{\boldsymbol{\psi}}$ and the precisions using Gibbs sampling.

Starting values for $\boldsymbol{t}$ and $\boldsymbol{u}$ are obtained using the following procedure. The component indicator $u_{ijr,1}$ is drawn uniformly from 1 to $U(1)$, and, if $y_{ijr} > 0$, the component indicator $u_{ijr,2}$ is drawn uniformly from 1 to $U(y_{ijr})$. To obtain a starting value for $t^{\star}_{ijr,2}$, we sample $t^{\star}_{ijr,2}$ from a $\text{Be}(y_{ijr}, 1)$-distribution. For $t^{\star}_{ijr,1} = \delta_{ijr,1}/\lambda'_{ijr}$, we sample $\delta_{ijr,1}$ from $\text{Exp}(\lambda'_{ijr})$ with $\lambda'_{ijr} = y_{ijr}$, if $y_{ijr} > 0$. If $y_{ijr} = 0$, λ'_{ijr} is set to a small value; we use $\lambda'_{ijr} = 0.1$.

Implementation

All MCMC programs were implemented in the low-level programming language `C` with proper incorporation of all sum-to-zero constraints. The library `GMRFLib` (Rue and Held, 2005, Appendix) was used in all programs. This library provides efficient numerical routines for sparse matrices to sample quickly from GMRFs. For uncorrelated multivariate APC models, both approaches, improved auxiliary mixture sampling (Frühwirth-Schnatter *et al.*, 2009) and a Metropolis-Hastings algorithm with a proposal for the linear predictor constructed using a second-order Taylor expansion of the log full conditional distribution, were used. For correlated multivariate APC models, the log likelihood was approximated using a second-order Taylor expansion and the resulting GMRF was used as a proposal in a Metropolis-Hastings algorithm for the linear predictor.

6.3 Integrated nested Laplace approximations

An alternative approach to MCMC for fully Bayesian inference in the class of latent Gaussian models has recently been proposed by Rue *et al.* (2009). Integrated nested Laplace approximations (INLA) are a deterministic technique which make MCMC sampling redundant by calculating exact approximations to the posterior marginal distributions. It has been shown that the quality of the approximations is very high for a wide range of applications (Rue *et al.*, 2009), see for example Martino *et al.* (2008) for the estimation of stochastic volatility models,

Martino *et al.* (2010) for survival analysis and Paul *et al.* (2010) for bivariate meta-analysis of diagnostic test studies. Rue *et al.* (2009) provide freely available software written in `C` which is easy to use under Linux, Windows and Macintosh via an `R`-Interface (R Development Core Team, 2010). The software can be downloaded from `www.r-inla.org`. We used INLA to analyse correlated multivariate APC models and APC models with covariates (Paper III and Paper IV). In the following we briefly present the main ideas of INLA.

A latent Gaussian model is a hierarchical model where the observations are non-Gaussian, but the latent field is Gaussian and only controlled by a few hyperparameters. In model (6) the latent field is $\boldsymbol{x} = (\boldsymbol{\mu}^\top, \boldsymbol{\theta}^\top, \tilde{\boldsymbol{\varphi}}^\top, \tilde{\boldsymbol{\psi}}^\top, \boldsymbol{z}^\top)^\top$, with $\boldsymbol{z} = (z_{111}, \ldots, z_{I11}, z_{121}, \ldots, z_{I21}, \ldots, z_{IJ1}, z_{112}, \ldots, z_{IJR})^\top$ and $\dim(\boldsymbol{x}) = R + I + J \times R + K \times R + I \times J \times R$. Assuming uncorrelated smoothing priors, uncorrelated overdispersion parameters and stratum-independent precision parameters, there are four hyperparameters $\boldsymbol{\theta}^\star = (\kappa_\theta, \kappa_\varphi, \kappa_\psi, \kappa_z)^\top$. Note, the notation conflict since the vector of hyperparameters in Rue *et al.* (2009) is denoted by $\boldsymbol{\theta}$, but we use $\boldsymbol{\theta}$ to denote the age effects. Hence we use $\boldsymbol{\theta}^\star$ to denote the vector of hyperparameters in INLA.

Assuming that the observations y_{ijr}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $r = 1, \ldots, R$ are conditionally independent the posterior distribution is given by

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}^\star | \boldsymbol{y}) \propto \pi(\boldsymbol{\theta}^\star)\pi(\boldsymbol{x}|\boldsymbol{\theta}^\star) \prod_{ijr} \pi(y_{ijr}|x_{ijr}, \boldsymbol{\theta}^\star).$$

Hence, the posterior marginals for each component x_{ijr}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $r = 1, \ldots, R$, of the latent field are given by

$$\pi(x_{ijr}|\boldsymbol{y}) = \int_{\boldsymbol{\theta}^\star} \underbrace{\int_{\boldsymbol{x}_{-ijr}} \pi(\boldsymbol{x}, \boldsymbol{\theta}^\star|\boldsymbol{y}) d\boldsymbol{x}_{-ijr}}_{\pi(x_{ijr}, \boldsymbol{\theta}^\star|\boldsymbol{y}) = \pi(x_{ijr}|\boldsymbol{\theta}^\star, \boldsymbol{y})\pi(\boldsymbol{\theta}^\star|\boldsymbol{y})} d\boldsymbol{\theta}^\star,$$

where $\boldsymbol{x}_{-ijr}$ denotes the vector $\boldsymbol{x}$ without component x_{ijr}. Since analytical integration of $\pi(\boldsymbol{x}, \boldsymbol{\theta}^\star|\boldsymbol{y})$ is usually not possible, MCMC techniques are the standard tool of choice. Instead of sampling, INLA directly approximates the posterior marginals with

$$\begin{aligned} \tilde{\pi}(x_{ijr}|\boldsymbol{y}) &= \int_{\boldsymbol{\theta}^\star} \tilde{\pi}(x_{ijr}|\boldsymbol{\theta}^\star, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta}^\star|\boldsymbol{y}) d\boldsymbol{\theta}^\star \\ &\approx \sum_u \tilde{\pi}(x_{ijr}|\boldsymbol{\theta}_u^\star, \boldsymbol{y}) \times \tilde{\pi}(\boldsymbol{\theta}_u^\star|\boldsymbol{y}) \times \Delta_u. \end{aligned}$$

Here, the approximation of $\pi(\boldsymbol{\theta}^\star|\boldsymbol{y})$ is a Laplace approximation, while for $\pi(x_{ijr}|\boldsymbol{\theta}^\star, \boldsymbol{y})$ three different choices are possible: a Gaussian, a Laplace and a simplified Laplace approximation. The default approximation is the simplified Laplace approximation, which is less time consuming than the full Laplace approximation and only slightly worse in terms of accuracy. In the last step posterior marginals for $\pi(x_{ijr}|\boldsymbol{y})$ are computed via numerical integration with area weights Δ_u. Posterior marginals for each precision parameter are computed in a similar way. For a detailed description of the approximations we refer to Rue *et al.* (2009).

6.4 Model choice

Usually the marginal likelihood is used for Bayesian model comparison. However, this requires that all prior distributions are proper because there is no unique scaling in the case of improper priors (Robert, 2001). Hence, we cannot use the marginal likelihood to compare different

uncorrelated multivariate APC models. However, if the models have the same second-stage-structure (for example joint age effects but stratum-specific period and cohort effects), and only differ by the choice of correlation structure for the priors, the marginal likelihood may be used (J.O. Berger, 2010, personal communication). In Paper III and IV we use the marginal likelihood computed by INLA to compare different correlated models.

Another frequently used model choice criterion is the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002). However, the use of DIC for models with many random effects has recently been criticised as complex models may be underpenalised (Plummer, 2008). Alternatively, proper scoring rules (Gneiting and Raftery, 2007) could be used: among them, mean ranked probability score, mean Dawid Sebastiani score (both available from MCMC) and the log score (from INLA). We used the mean ranked probability score and the mean Dawid Sebastiani score in Paper I, and their multivariate analogues in Paper III + IV. In Paper IV we additionally use the log score. The log score is computed as $1/(IJR)\sum_{i,j,r}\log(\text{CPO}_{ijr})$ where CPO_{ijr} denotes the conditional predictive ordinate (Pettit, 1990). In both applications of Paper III a large number of CPO values, returned by INLA, are classified as unreliable. INLA indicates such failures, so that the desired leave-one-out quantities could be computed "manually" (Held *et al.*, 2010). However, due to the large number of failures we decided not to compute the log score for these applications.

Thesis Summary

This thesis consists of four papers, presented in chronological order. We first briefly summarise the contents of each paper:

Paper I

In Paper I, **The analysis of heterogeneous time trends in multivariate age-period-cohort models** by Andrea Riebler and Leonhard Held, a novel Bayesian approach to estimate relative risk in multivariate APC models is proposed. Model choice is performed using cross-validation and proper scoring rules (Marshall and Spiegelhalter, 2003; Gneiting and Raftery, 2007; Czado *et al.*, 2009). The methodology is applied to data on COPD mortality of males in England and Wales (Hansell *et al.*, 2003; Hansell, 2004) and on overall mortality rates of women in Denmark and Norway (Jacobsen *et al.*, 2004). A comparison between Bayesian estimates and those obtained from a classical maximum-likelihood analysis is given.

This work was inspired by the PhD thesis of Hansell (2004) which was partly supervised by L. Held. As part of this supervision, L. Held implemented an `R`-function for the analysis of multivariate APC models using classical maximum-likelihood inference which was the starting point for this paper. Together, we extended the Bayesian univariate APC model to the multivariate case. I implemented the MCMC routines in the low level programming language `C` using the `GMRFLib` library (Rue and Held, 2005, Appendix) for fast and exact simulation from GMRFs. Two different approaches were considered: Auxiliary mixture sampling (Frühwirth-Schnatter *et al.*, 2009) and a Metropolis-Hastings algorithm with a proposal constructed using a second-order Taylor expansion of the log likelihood (Rue and Held, 2005). A first draft of the methodology was written by L. Held. I commented on this original draft, conducted all analyses and wrote a first draft of the application sections. Together, we finalised the paper.

The main contribution of the paper is on tackling the well-known non-identifiability problems in APC models (Clayton and Schifflers, 1987b; Besag *et al.*, 1995). Moreover, an additional non-identifiability problem for data not observed on equal time intervals (Holford, 2006) could be avoided through the use of smoothing priors.

Paper II

In Paper II, **A conditional approach for inference in multivariate age-period-cohort models** by Leonhard Held and Andrea Riebler, an alternative frequentist approach for inference in multivariate APC models is derived. Within a conditional framework, differences in stratum-specific time effects are modelled directly, while ignoring a number of nuisance parameters present in the original formulation. A polychotomous regression approach is proposed based on a multinomial logistic regression model. The usage of smoothing splines is suggested to stabilise estimates of relative time trends, if necessary.

This work is based on Paper I. L. Held proposed and developed the conditional multinomial model formulation and wrote a first draft. I extended the theoretical part and compared software for the analysis of multinomial logistic regression models that additionally provides the possibility for parameter smoothing and adjustments for overdispersion. Using the `R`-package `VGAM` (Yee, 2009) I performed all analyses including the exact computation of point-wise confidence intervals for the relative risk parameters. Together, we finalised the paper.

The main contribution of the paper is that the suggested conditional approach simplifies the

analysis and interpretation of multivariate APC models. Parameters of interest, namely the relative risks, are directly modelled and standard software for multinomial logistic regression can be used for model estimation. Furthermore, the approach can be used to analyse age-specific rates that are stratified by more than one variable. Additionally, unmeasured confounding can be handled to some extent, so that relative risks may be estimated with higher precision.

Paper III

In Paper III, **Correlated multivariate age-period-cohort models** by Andrea Riebler, Leonhard Held and Håvard Rue, an extended Bayesian approach for correlated multivariate APC models is presented. The use of correlated smoothing priors and correlated overdispersion parameters is proposed to capture the dependence that may exist between multiple outcomes. Algorithmic routines are implemented using MCMC and INLA (Rue *et al.*, 2009) based on a uniform correlation structure. Two applications are presented. The results obtained by an ordinary multivariate APC model are compared with those obtained by the correlated model formulation using DIC (Spiegelhalter *et al.*, 2002), proper scoring rules (Gneiting and Raftery, 2007) and the log marginal likelihood.

This work extends Paper I. L. Held proposed the use of correlated smoothing priors and correlated overdispersion parameters within multivariate APC models. After discussion of the general formulation with L. Held, I worked out most of the details of the correlated multivariate APC model. During a research visit in the group of H. Rue in Trondheim, funded by the Research Council of Norway, I assisted H. Rue in integrating the uniform correlation structure into INLA. In addition, I implemented the MCMC routines in the low level programming language `C` using the `GMRFLib` library (Rue and Held, 2005, Appendix). I wrote a draft of the manuscript. All authors then commented on the draft, which I finalised.

The main contribution of the paper is that the correlated formulation can improve the precision of relative risks and is useful for projection, as exemplified in the two applications of the paper. The correlated approach involves a Kronecker product precision matrix. As consequence of integrating this structure into INLA a wide range of latent GMRF models can be correlated as components of more general structured additive regression models.

Appendix I presents an early version of Paper III as it will appear in the proceedings of the 25th International Workshop on Statistical Modelling in Glasgow $5-9$th July 2010. Results on projecting incidence or mortality rates are not included in this version.

Appendix II presents the full conditional distributions derived for the ordinary and correlated multivariate APC model.

Appendix III presents a program description for the MCMC program. With this program, ordinary and different formulations of correlated multivariate APC models can be analysed. In addition, sets of parameters can be excluded from the analysis, leading to e.g. an age-cohort model if period effects are excluded.

Paper IV

In Paper IV, **Suicide mortality in Switzerland: Gender-specific differences and the impact of family integration** by Andrea Riebler, Leonhard Held, Håvard Rue and Matthias Bopp, ordinary and correlated multivariate APC models are applied to suicide mortality of males and females in Switzerland. First, heterogeneous time trends between males and females are analysed. A model with correlated gender-specific age and cohort effects and correlated

overdispersion parameters is classified as the best model. Elderly and young men are observed to have a three-fold higher risk to commit suicide than their female peers in the same age group. Further the impact of family integration on suicide mortality is investigated. Amongst other model formulations, the joint period effects are replaced by correlated covariate effects. We found that a high family integration has a decreasing effect on suicide risk for both sexes. However, in terms of model choice criteria, the effect of the covariate is not strong enough to completely replace the general period effects, so that a correlated multivariate APC model without covariates is still preferred.

This work applies the methodology developed in Paper I and III and strongly depends on the integration of the uniform correlation structure into INLA performed during a research visit to the group of H. Rue. M. Bopp proposed analysing gender-specific suicide rates in Switzerland with multivariate APC models and provided all data. I performed an extensive literature search on the impact of different covariates on suicide risk. The effect of marital status and family integration are widely discussed. Hence, I proposed examining their effect on Swiss suicide data in the context of APC models. After discussion with L. Held, I conducted all analyses in the paper as well as the writing of the manuscript. M. Bopp and L. Held commented on the manuscript.

The main contribution of this paper is to illustrate the use and applicability of (correlated) multivariate APC models. The analysis revealed strong gender-specific differences in suicide mortality. Similar risk factors may act on age and overdispersion whereas we found no strong correlation between cohort effects. Further, we found an impact of family integration on Swiss suicide risk. In contrast to standard time-series models, the models we proposed for integrating covariate information keep the age-specific structure of the data, so that more information can be gained.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle, *in* B. N. Petrov and F. Csaki (eds), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, pp. 267–281.

Baker, A. and Bray, I. (2005). Bayesian projections: What are the effects of excluding data from younger age groups?, *American Journal of Epidemiology* **162**: 798–805.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, John Wiley & Sons, Ltd, Chichester.

Berzuini, C. and Clayton, D. (1994). Bayesian analysis of survival on multiple time scales, *Statistics in Medicine* **13**: 823–838.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation of stochastic-systems, *Statistical Science* **10**: 3–41.

Bray, I. (2002). Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality, *Journal of the Royal Statistical Society - Series C* **51**: 151–164.

Bray, I., Brennan, P. and Boffetta, P. (2001). Recent trends and future projections of lymphoid neoplasms - a Bayesian age-period-cohort analysis, *Cancer Causes & Control* **12**: 813–820.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models, *Applied Statistics* **33**: 38–44.

Brown, C. C. and Kessler, L. G. (1988). Projections of lung cancer mortality in the United States: 1985-2025, *Journal of the National Cancer Institute* **80**: 43–51.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn, Springer, New York.

Clayton, D. and Schifflers, E. (1987a). Models for temporal variation in cancer rates. I: Age-period and age-cohort models, *Statistics in Medicine* **6**: 449–467.

Clayton, D. and Schifflers, E. (1987b). Models for temporal variation in cancer rates. II: Age-period-cohort models, *Statistics in Medicine* **6**: 469–481.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data, *Biometrics* **65**: 1254–1261.

Davison, A. C. (2003). *Statistical Models*, Cambridge University Press, Cambridge.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors, *Journal of the Royal Statistical Society - Series C* **50**: 201–220.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models*, 2nd edn, Springer, Berlin.

Fienberg, S. E. and Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology* **10**: 1–67.

Fisher, R. A. (1958). *Statistical Methods for Research Workers*, 13th ed.(rev.) edn, Oliver & Boyd, Edinburgh.

Frühwirth-Schnatter, S. and Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models, *Computational Statistics & Data Analysis* **51**: 3509–3528.

Frühwirth-Schnatter, S., Frühwirth, R., Held, L. and Rue, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data, *Statistics and Computing* **19**: 479–492.

Fu, W. (2000). Ridge estimator in singular design with application to age-period-cohort analysis of disease rates, *Communications in Statistics - Theory and Methods* **29**: 263–278.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models, *Statistics and Computing* **7**: 57–68.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* **102**: 359–378.

Grimmett, G. R. and Stirzaker, D. R. (2001). *Probability and Random Processes*, 3rd edn, Oxford University Press, New York.

Hansell, A., Knorr-Held, L., Best, N., Schmid, V. and Aylin, P. (2003). COPD mortality trends 1950-1999 in England & Wales - did the 1956 Clean Air Act make a detectable difference?, *Epidemiology* **14**: S55–S55.

Hansell, A. L. (2004). *The Epidemiology of Chronic Obstructive Pulmonary Disease in the UK: Spatial and Temporal Variations*, PhD thesis, Faculty of Medicine of the University of London, Imperial College, St Mary's Campus.

Harvey, A. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*, reprinted edn, Cambridge University Press, Cambridge.

Held, L. and Riebler, A. (2010). A conditional approach for inference in multivariate age-period-cohort models, *Statistical Methods in Medical Research*. Revised version under review.

Held, L., Schrödle, B. and Rue, H. (2010). Posterior cross-validatory predictive checks: A comparison of MCMC and INLA, *in* T. Kneib and G. Tutz (eds), *Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir*, Physica-Verlag.

Heuer, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines, *Biometrics* **53**: 161–177.

Hilbe, J. M. (2007). *Negative Binomial Regression*, Cambridge University Press, Cambridge.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, *Journal of the American Statistical Association* **91**: 1461–1473.

Holford, T. (2005). Age-period-cohort analysis, *in* P. Armitage and T. Colton (eds), *Encyclopaedia of Biostatistics*, 2nd edn, John Wiley and Sons, West Sussex, pp. 105–123.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates, *Biometrics* **39**: 311–324.

Holford, T. R. (1991). Understanding the effects of age, period and cohort on incidence and mortality rates, *Annual Reviews of Public Health* **12**: 425–457.

Holford, T. R. (1992). Analysing the temporal effects of age, period and cohort, *Statistical Methods in Medical Research* **1**: 317–337.

Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals, *Statistics in Medicine* **25**: 977–993.

Jacobsen, R., von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different?, *European Journal of Epidemiology* **19**: 117–121.

Keiding, N. (1990). Statistical inference in the Lexis diagram, *Philosophical Transactions: Physical Sciences and Engineering* **332**: 487–509.

Knorr-Held, L. and Rainer, E. (2001). Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction, *Biostatistics* **2**: 109–129.

Kuang, D., Nielsen, B. and Nielsen, J. (2008). Identification of the age-period-cohort model and the extended chain-ladder model, *Biometrika* **95**: 979–986.

Künsch, H. R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice, *Biometrika* **74**: 517–524.

Lexis, W. (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*, Karl Trübner, Straßburg.

Marshall, E. C. and Spiegelhalter, D. J. (2003). Approximate cross-validatory predictive checks in disease mapping models, *Statistics in Medicine* **22**: 1649–1660.

Martino, S., Aas, K., O, L., Neef, L. R. and Rue, H. (2008). Estimating stochastic volatility models using integrated nested Laplace approximations, *Technical report*, Norwegian University of Science and Technology, Trondheim. Available from: `http://www.math.ntnu.no/~hrue/r-inla.org/papers/stochvol-nr.pdf`.

Martino, S., Akerkar, R. and Rue, H. (2010). Bayesian inference for survival models, *Technical report*, Norwegian University of Science and Technology, Trondheim. Available from: `http://www.math.ntnu.no/~hrue/r-inla.org/papers/Tech-Report-02-10.pdf`.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Chapman and Hall, London.

Osmond, C. (1985). Using age, period and cohort models to estimate future mortality rates, *International Journal of Epidemiology* **14**: 124–129.

Osmond, C. and Gardner, M. J. (1982). Age, period and cohort models applied to cancer mortality rates, *Statistics in Medicine* **1**: 245–259.

Osmond, C. and Gardner, M. J. (1989). Age, period, and cohort models. Non-overlapping cohorts don't resolve the identification problem, *American Journal of Epidemiology* **129**: 31–35.

Paul, M., Riebler, A., Bachmann, L. M., Rue, H. and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations, *Statistics in Medicine* **29**: 1325–1339.

Pettit, L. I. (1990). The conditional predictive ordinate for the normal-distribution, *Journal of the Royal Statistical Society - Series B* **52**: 175–184.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison, *Biostatistics* **9**: 523–539.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: `http://www.R-project.org`.

Riebler, A. and Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models, *Biostatistics* **11**: 57–69.

Riebler, A., Held, L. and Rue, H. (2010a). Correlated multivariate age-period-cohort models, *Technical report*, University of Zurich.

Riebler, A., Held, L., Rue, H. and Bopp, M. (2010b). Suicide mortality in Switzerland: Gender-specific differences and the impact of family integration, *Technical report*, University of Zurich.

Robert, C. P. (2001). *The Bayesian Choice*, 2nd edn, Springer, New York.

Robertson, C. and Boyle, P. (1986). Age, period and cohort models: the use of individual records, *Statistics in Medicine* **5**: 527–538.

Robertson, C., Gandini, S. and Boyle, P. (1999). Age-period-cohort models: a comparative study of available methodologies, *Journal of Clinical Epidemiology* **52**: 569–583.

Rodgers, W. L. (1982). Estimable functions of age, period, and cohort effects, *American Sociological Review* **47**: 774–787.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC Press, London.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society - Series B* **71**: 319–392.

Schmid, V. J. and Held, L. (2007). Bayesian age-period-cohort modeling and prediction - BAMP, *Journal of Statistical Software* **21**: 1–15.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R. and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society - Series B* **64**: 583–616.

Thygesen, L. C., Knudsen, L. B. and Keiding, N. (2005). Modelling regional variation of first-time births in Denmark 1980-1994 by an age-period-cohort model, *Demographic Research* **13**: 573–596.

Yang, Y., Fu, W. J. J. and Land, K. C. (2004). A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models, *Sociological Methodology* **34**: 75–110.

Yang, Y., Schulhofer-Wohl, S., Fu, W. and Land, K. (2008). The intrinsic estimator for age-period-cohort analysis: What it is and how to use it, *American Journal of Sociology* **113**: 1697–1736.

Yee, T. W. (2009). *A User's Guide to the* `VGAM` *Package*. Available from: `http://www.stat.auckland.ac.nz/~yee/VGAM`.

Zheng, T. Z., Holford, T. R., Chen, Y. T., Ma, J. Z., Flannery, J., Liu, W., Russi, M. and Boyle, P. (1996). Time trend and age-period-cohort effect on incidence of thyroid cancer in Connecticut, 1935-1992, *International Journal of Cancer* **67**: 504–509.

PAPER I

The analysis of heterogeneous time trends in multivariate age-period-cohort models

Andrea Riebler & Leonhard Held

Paper published in *Biostatistics* (2010), **11**, 1, pp. 57–69.

Biostatistics (2010), **11**, 1, *pp*. 57–69
doi:10.1093/biostatistics/kxp037
Advance Access publication on October 13, 2009

R

The analysis of heterogeneous time trends in multivariate age–period–cohort models

ANDREA RIEBLER, LEONHARD HELD*

Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland
leonhard.held@ifspm.uzh.ch

SUMMARY

Age–period–cohort (APC) models are frequently used to analyze mortality or morbidity rates stratified by age group and period. For the case in which rates are given in different strata, multivariate APC models have been considered only recently. Such models share a set of parameters, for example, the age effects, while the other parameters may vary across strata. We show that differences of strata-specific effects are identifiable. We then propose a Bayesian approach based on smoothing priors to estimate multivariate APC models. This provides an alternative to maximum likelihood (ML) estimates of relative risk in the case of equal intervals and gives useful results even in the case of unequal intervals, where ML estimates have severe artifacts. This is illustrated with data on female mortality in Denmark and Norway and data on chronic obstructive pulmonary disease mortality of males in England and Wales, stratified by 3 different areas: Greater London, conurbations excluding Greater London, and nonconurbation areas.

Keywords: Heterogeneous time trends; Identifiability; Multivariate age–period–cohort model; Overdispersion; Relative risk; Smoothing.

1. INTRODUCTION

Age–period–cohort (APC) models are frequently used to analyze mortality or morbidity rates stratified by age group and period (Holford, 1983, 1998; Clayton and Schifflers, 1987). Bayesian versions of the APC model have been proposed in Berzuini *and others* (1993), Berzuini and Clayton (1994), and Besag *and others* (1995). Bayesian formulations assume some sort of smoothness of age, period, and cohort effects and are particularly useful for predicting future rates (Knorr-Held and Rainer, 2001; Bray, 2002) as the prior models on period and cohort effects can be stochastically extrapolated to future values. Additional adjustments for overdispersion are easily incorporated.

Multivariate APC models to capture heterogeneous time trends in different strata have been considered only recently (Hansell *and others*, 2003; Hansell, 2004; Jacobsen *and others*, 2004). Such models share a set of parameters, for example, the age effects, while the remaining parameters can be different across strata. Multivariate APC models can be used to analyze heterogeneous time trends in different geographical areas, for different diseases, or simply by gender. Despite the well-known nonidentifiability of

*To whom correspondence should be addressed.

linear time trends in APC models, we show that differences of period or cohort effects between strata are identifiable and can be interpreted as log relative risks.

Up to now, statistical inference has used Poisson regression models, but this can lead to unstable maximum likelihood (ML) estimates. In this paper, we propose a Bayesian approach to estimate multivariate APC models based on smoothing priors. This provides an alternative to ML estimates of relative risk in the case of equal intervals and enables us to estimate relative risks even in the case of unequal intervals, where the cyclical pattern of unregularized ML estimates causes severe artifacts in the relative risk estimates.

In Section 2, we describe our approach in the case of equal intervals, with some specific details on statistical inference via Markov chain Monte Carlo (MCMC) in Section 2.3. Section 2.4 describes how model assessment is performed. Section 2.5 compares the Bayesian estimates with those obtained from a classical ML analysis for the model considered best in an analysis of mortality for women in Norway and Denmark aged 0–84 years during the period 1960–1999 (Jacobsen *and others*, 2004). In Section 3, we turn to the case of unequal intervals, where Bayesian smoothing priors enable us to estimate relative risks even in cases where this is impossible with ML. Problems with ML estimates in the univariate case have been recognized recently (Holford, 2006) and extend to the multivariate case, as outlined in Section 3.1. We provide a detailed analysis of yearly data on chronic obstructive pulmonary disease (COPD) mortality in England and Wales, 1950–1999. Data were obtained by 10-year age bands and are stratified by 3 different areas: Greater London, conurbations excluding Greater London, and nonconurbation areas in Section 3.2. We end with some discussion in Section 4.

2. MULTIVARIATE APC MODELS WITH EQUAL TIME INTERVALS

2.1 *The univariate APC model*

To begin, let n_{ij} denote the number of persons at risk in age group i ($i = 1, \ldots, I$) and period j ($j = 1, \ldots, J$). We assume that the number of cases y_{ij} in age group i during period j has a Poisson distribution with rate $n_{ij}\lambda_{ij}$ and that the likelihood for the entire data is the corresponding product of the Poisson terms.

If the age group intervals are of the same width as the period intervals, the standard APC model (e.g. Clayton and Schifflers, 1987) could be adopted. This decomposes the log relative rate $\eta_{ij} = \log\{\lambda_{ij}\}$ additively into an overall level μ, age effects θ_i, period effects φ_j, and cohort effects ψ_k:

$$\eta_{ij} = \mu + \theta_i + \varphi_j + \psi_k, \tag{2.1}$$

where the cohort index $k = 1, \ldots, K$ is given by $k = I - i + j$. To assure identifiability, 2 types of additional constraints are necessary. First, restrictions have to be imposed on each block of parameters $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, and $\boldsymbol{\psi}$ to make the overall mean μ identifiable. A typical choice is $\sum_i \theta_i = 0$, $\sum_j \varphi_j = 0$, and $\sum_k \psi_k = 0$. The second redundancy is due to the linear dependence of the cohort index k on i and j since, for any value of a, the linear transformations

$$\theta_i \to \theta_i + a\left(i - \frac{I+1}{2}\right), \quad \varphi_j \to \varphi_j - a\left(j - \frac{J+1}{2}\right), \quad \text{and} \quad \psi_k \to \psi_k + a\left(k - \frac{K+1}{2}\right)$$

will still fulfil $\sum_i \theta_i = 0$, $\sum_j \varphi_j = 0$, and $\sum_k \psi_k = 0$ but leave η_{ij} unchanged for all i and j. Hence, only nonlinear trends of the age, period, or cohort blocks are interpretable but linear trends are not. This problem is well known and thoroughly discussed in the literature (Holford, 1983; Clayton and Schifflers, 1987; Berzuini *and others*, 1993).

2.2 *The multivariate APC model*

Suppose now that the data are stratified further, for example, by different geographical areas, different causes of mortality, or simply by gender. Let y_{ijr} denote the count in age group i, period j, and stratum $r = 1, \ldots, R$ with associated population count n_{ijr}. Again assume that the y_{ijr} are Poisson with mean $n_{ijr}\lambda_{ijr}$. A specific multivariate APC model, potentially useful in many applications, assumes that the age effects are equal, whereas period and cohort effects may differ across strata. The overall level may vary across strata, which defines the linear predictor $\eta_{ijr} = \log\{\lambda_{ijr}\}$ as

$$\eta_{ijr} = \mu_r + \theta_i + \varphi_{jr} + \psi_{kr}. \tag{2.2}$$

It is obvious how to modify this model in the case of unequal period but equal cohort effects or vice versa. Similarly, the age effects may vary across strata but the period effects may be fixed.

As before, sum-to-zero constraints have to be imposed: $\sum_i \theta_i = 0$, $\sum_j \varphi_{jr} = 0$ for all $r = 1, \ldots, R$, and $\sum_k \psi_{kr} = 0$ for all $r = 1, \ldots, R$. However, the additional nonidentifiability due to the linear dependence between age, period, and cohort effects is still present. Indeed, for any value of a, the linear transformations

$$\begin{aligned}
\theta_i &\to \theta_i + a\left(i - \frac{I+1}{2}\right), \\
\varphi_{jr} &\to \varphi_{jr} - a\left(j - \frac{J+1}{2}\right) \quad \text{for } r = 1, \ldots, R, \quad \text{and} \\
\psi_{kr} &\to \psi_{kr} + a\left(k - \frac{K+1}{2}\right) \quad \text{for } r = 1, \ldots, R
\end{aligned} \tag{2.3}$$

will still fulfil all sum-to-zero constraints and leave the linear predictor η_{ijr} unchanged for all i, j, and r.

Consider now the difference $\Delta_j = \varphi_{j,r_1} - \varphi_{j,r_2}$ of 2 period effects for 2 arbitrary strata r_1 and r_2. From (2.3) it is easy to see that Δ_j is left unchanged for any value of a. Therefore, despite the nonidentifiability of the period effects, differences between period effects are identifiable, as long as the strata share the same age effects. Differences $\Delta_k = \psi_{k,r_1} - \psi_{k,r_2}$ of 2 cohort effects are also identifiable. However, the differences are not identifiable if all 3 effects are allowed to vary as a will then depend on stratum r.

Let $\Delta_\mu = \mu_{r_1} - \mu_{r_2}$. In the case where only the period effects differ across strata, the adjusted difference $\tilde{\Delta}_j = \Delta_\mu + \Delta_j$ can be interpreted as the "log relative risk" in period j and stratum r_1, relative to stratum r_2. Note that the age effects and cohort effects are no longer present. Since they are assumed to be the same across strata, they cancel out in the difference of the log relative rates of strata r_1 and r_2. Similarly, if the cohort effects differ across strata, $\tilde{\Delta}_k = \Delta_\mu + \Delta_k$ is the log relative risk of cohort k in stratum r_1, relative to stratum r_2. If both period and cohort effects are allowed to vary across strata, the log relative risk $\tilde{\Delta}_{jk} = \Delta_\mu + \Delta_j + \Delta_k$ depends both on period j and on cohort k. In this case,

$$\tilde{\Delta}_j = \frac{1}{K}\sum_k \tilde{\Delta}_{jk} \quad \text{and} \quad \tilde{\Delta}_k = \frac{1}{J}\sum_j \tilde{\Delta}_{jk}$$

can be interpreted as "average log relative risk" due to $\sum_k \Delta_k = 0$ and $\sum_j \Delta_j = 0$, respectively. Therefore, $\exp(\tilde{\Delta}_j)$ is the geometrically averaged relative risk in period j and likewise $\exp(\tilde{\Delta}_k)$ in cohort k.

We now describe 2 options for statistical inference. A standard approach would treat all age, period, and cohort effects as factors and would employ standard Poisson regression procedures for estimating the unknown parameters by ML (Jacobsen *and others*, 2004). The nonidentifiability problem (2.3) is usually dealt with setting an additional parameter equal to zero.

However, certain complications arise. In particular, the fit in the cells ($i = 1$, $j = J, r = 1, \ldots, R$) must be perfect since the cohort parameters $\psi_{K,r}$, $r = 1, \ldots, R$, enter in one and only one of these cells. The same holds for the cells ($i = I$, $j = 1, r = 1, \ldots, R$) and the cohort parameters $\psi_{1,r}$, $r = 1, \ldots, R$. The residuals in these cells will therefore always be zero. Further problems arise if the number of cases associated with certain cohort parameters are zero, in which case the iterative algorithms to compute the ML estimates will diverge. It is best to remove such parameters in advance.

A Bayesian approach avoids these problems through employing smoothing priors on age, period, and cohort effects, as described in detail in Besag *and others* (1995) and Knorr-Held and Rainer (2001). For example, consider the period effects $\boldsymbol{\varphi}_r = (\varphi_{1,r}, \ldots, \varphi_{J,r})$ for a specific stratum r. A smoothing prior based on second differences can be written as

$$p(\boldsymbol{\varphi}_r|\kappa_r) \propto \exp\left(-\frac{\kappa_r}{2}\sum_{j=3}^{J}(\varphi_{j,r} - 2\varphi_{j-1,r} + \varphi_{j-2,r})^2\right), \tag{2.4}$$

where the precision parameter κ_r, which determines the amount of smoothing, is treated as unknown and also estimated from the data. The equivalent directed formulation is called a "second-order random walk" (RW2) defined by $\varphi_{j,r} \sim \mathcal{N}(2\varphi_{j-1,r} - \varphi_{j-2,r}, \kappa_r^{-1})$, $j = 3, \ldots, J$, with (improper) independent uniform priors for both $\varphi_{1,r}$ and $\varphi_{2,r}$. This model is the discrete-time analog of a cubic smoothing spline (Fahrmeir and Knorr-Held, 2000; Rue and Held, 2005). The same prior is used for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)$ and $\boldsymbol{\psi}_r = (\psi_{1,r}, \ldots, \psi_{K,r})$ in (2.2).

This prior is a natural choice as it penalizes deviations from a linear trend but does not resolve the nonidentifiability problem described earlier. However, in a Bayesian framework it is not crucial to ensure identifiability of latent parameters as long as the linear predictors η_{ijr} are identifiable; for details see the discussion following Besag *and others* (1995). We therefore use no additional constraint except the standard sum-to-zero restrictions.

Adjustments for overdispersion can be incorporated directly in the model by introducing additional independent Gaussian variables $z_{ijr} \sim \mathcal{N}(0, \delta^{-1})$ in the linear predictor (2.2):

$$\eta_{ijr} = \mu_r + \theta_i + \varphi_{jr} + \psi_{kr} + z_{ijr}. \tag{2.5}$$

In the same way as for the other hyperparameters, we assign a gamma prior to the hyperparameter δ. Alternatively, independent random effects may be added for each cohort k and stratum r, as a referee has suggested.

2.3 *Inference by MCMC*

Following Besag *and others* (1995), we reparameterized the model from z_{ijr} to η_{ijr} to obtain multivariate Normal full conditional distributions for $\boldsymbol{\theta}$, $\boldsymbol{\varphi}_r$, and $\boldsymbol{\psi}_r$, $r = 1, \ldots, R$. Block updating of these main effects allows proper incorporation of sum-to-zero restrictions as described in Knorr-Held and Rue (2002); see also Rue and Held (2005) for algorithmic details. The intercepts μ_r and all hyperparameters can also be updated by Gibbs sampling.

We considered 2 possibilities to update the linear predictor η_{ijr}, which has a nonstandard full conditional distribution. The first uses univariate Metropolis–Hastings updates with a Gaussian proposal distribution based on a second-order Taylor expansion of the log full conditional distribution (Rue and Held, 2005, Section 4.4). The second applies the auxiliary mixture sampling approach proposed by Frühwirth-Schnatter and Wagner (2006) and further developed by Frühwirth-Schnatter *and others* (2009). This method introduces 4 additional auxiliary variables for each observation y_{ijr} to obtain a Gaussian full conditional distribution for each η_{ijr}, so that Gibbs sampling becomes possible. While the 2 methods yield identical results, the Metropolis–Hastings update turned out to be more efficient in terms of the

relative effective sample size (ESS) (Kass *and others*, 1998). We therefore chose this option in the following. Both algorithms have been implemented in C using the library GMRFLib (Rue and Held, 2005, Appendix B). Both the source code and the executable of the Metropolis–Hastings program, accompanied by a program description and the data set analyzed in Section 2.5, can be found as supplementary material, available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org).

As in Knorr-Held and Rainer (2001), we use highly dispersed prior distributions, namely Ga(1, 0.000, 05), for all hyperparameters associated with smoothing priors of the form (2.4) and a Ga(1, 0.005) prior distribution for δ. Locally uniform priors are assumed for the intercepts μ_r, $r = 1, \ldots, R$. All results presented are based on $N = 5000$ samples, collected by saving every 20th iteration after a burn-in interval of 20 000 iterations.

Both pointwise and simultaneous credible bands (Besag *and others*, 1995, p. 30) have been computed for the relative risks. Simultaneous credible bands allow for an additional check whether a constant relative risk is plausible. The computation of confidence intervals in the ML approach is not as obvious due to the additional constraints implied by (2.3). For example, Jacobsen *and others* (2004) only report point estimates of relative risk.

We have routinely examined convergence and mixing diagnostics. We have calculated the ESS and visually checked the corresponding trace plots. Nearly all η_{ijr} turned out to be virtually independent with ESS very close to the nominal sample size in the following application. Because of the nonidentifiability (2.3), convergence diagnostics cannot be applied to the age, period, or cohort effects.

2.4 *Model choice*

For Bayesian model comparison, a frequently used method is the deviance information criterion (DIC) (Spiegelhalter *and others*, 2002). However, in hierarchical models with many random effects, such as model (2.5), DIC tends to underpenalize complex models. Cross-validatory predictive checks might be more appropriate in this case. Unfortunately, full leave-one-out cross-validatory approaches can be very time-consuming in MCMC as the full analysis has to be repeated removing each data point in turn (Plummer, 2008).

Alternatives are based on running a single MCMC analysis. Importance sampling methods, for example, replicate the output in a way that theoretically removes the influence of a single observation. In posterior predictive model checking, the random effects are estimated from the full data and a new replicate observation is generated from its conditional distribution. However, in this case the random effects depend directly on the observed data so that the actually observed data point has strong influence on its replicate. The consequence is that they will tend to agree too well. Marshall and Spiegelhalter (2003) propose to lessen the influence of the observed data point by replicating both random effects and data at each iteration without repeatedly refitting the model with individual observations removed. This is particularly attractive in our setting, where we have a random effect η_{ijr} for each observation y_{ijr}. For example, in model (2.5) we generate a new replicate η_{ijr}^{rep} from $\mathcal{N}(\mu_r + \theta_i + \varphi_{jr} + \psi_{kr}, \delta^{-1})$ followed by a simulation of y_{ijr}^{rep} from a Poisson distribution with mean $n_{ijr} \exp(\eta_{ijr}^{\text{rep}})$. Here, the conservatism introduced is moderate since y_{ijr} only influences η_{ijr}^{rep} indirectly through $\mu_r, \theta_i, \varphi_{jr}, \psi_{kr}$, and δ^{-1}. Marshall and Spiegelhalter (2003) showed that this approach provides a better approximation to cross-validation than either importance sampling or posterior predictive alternatives.

The replicated data points can then be used to provide sound model choice criteria based on proper scoring rules (Gneiting and Raftery, 2007). For count data, Czado *and others* (2009) proposed the mean ranked probability score

$$\overline{\text{RPS}} = \frac{1}{I \cdot J \cdot R} \sum_{i,j,r} \left(\frac{1}{N} \sum_{n=1}^{N} \left| y_{ijr_{(n)}}^{\text{rep}} - y_{ijr} \right| - \frac{1}{N} \sum_{n=1}^{N/2} \left| y_{ijr_{(n)}}^{\text{rep}} - y_{ijr_{(n+N/2)}}^{\text{rep}} \right| \right),$$

where $|\cdot|$ denotes the absolute value and $y^{\text{rep}}_{ijr_{(n)}}$ the nth replicate for observation y_{ijr}, or the mean Dawid–Sebastiani score

$$\overline{\text{DSS}} = \frac{1}{I \cdot J \cdot R} \sum_{i,j,r} \left[\left(\frac{y_{ijr} - \overline{y^{\text{rep}}_{ijr}}}{\sigma_{y^{\text{rep}}_{ijr}}} \right)^2 + 2 \log \sigma_{y^{\text{rep}}_{ijr}} \right],$$

where $\overline{y^{\text{rep}}_{ijr}}$ and $\sigma_{y^{\text{rep}}_{ijr}}$ denote mean and standard deviation of the replicates $y^{\text{rep}}_{ijr_{(n)}}$, $n = 1, \ldots, N$. Smaller values of either score are to be preferred.

In the following, uppercase letters for age (A), period (P), or cohort (C) denote a model where the corresponding effect is assumed to be the same over the different strata, while lowercase letters (a, p, or c) denote a model where the effect is allowed to differ across strata.

2.5 *Application: mortality of Danish and Norwegian women*

We reanalyzed data on overall mortality, aggregated to 5-year age group and period intervals, for all Danish and Norwegian women aged 0–84 years during the period 1960–1999 (Jacobsen *and others*, 2004). Following Baker and Bray (2005), we also include data on the younger age groups from 0 to 39 years. The analysis of mortality in Danish women was put up for lively discussion in the press when Kesteloot (2001) speculated that the increase in mortality could be linked to a role model effect of Queen Margrethe II of Denmark, who has been a known cigarette smoker. If there is a link, we would expect a pronounced period effect around her ascension to the throne in 1972.

However, according to our model choice criteria, a model with separate age and cohort effects but joint period effects (model aPc) was preferred, compare Table 1. Quite remarkably, the mean scores $\overline{\text{RPS}}$ and $\overline{\text{DSS}}$ were even smaller than those in the apc model, which does not allow estimation of relative risks. This indicates that there is no evidence of different period effects as suspected by Kesteloot (2001). Relative risk estimates of Danish compared to Norwegian women based on ML and the Bayesian approach from the aPc model are displayed in Figure 1. As indicated by small tick marks on the x-axis, the ML approach did not provide estimates for all cohorts. When computing the average relative risk $\exp(\tilde{\Delta}_i)$, the missing values had to be linearly extrapolated to provide results comparable with the Bayesian approach. Otherwise, the 2 methods give fairly similar estimates.

The estimated overdispersion parameters obtained from the ML analysis range from 4.83 to 64.26, depending on the chosen model, reflecting substantial overdispersion. Adjustments for overdispersion are obviously necessary.

Note that Jacobsen *and others* (2004) have assumed joint age effects, whereas our analysis has identified clear evidence for separate age effects. This may be caused by the inclusion of the age groups below 40 years (see Figure 1). However, note that the average relative risk of Danish compared to Norwegian women not only increases in younger age groups but also decreases in older age groups. This may be due to different causes of death in the 2 countries, with Denmark having a larger proportion of cancer deaths (Helweg-Larsen *and others*, 1998) where most cases occur in the mid-age groups (40–65/70)

Table 1. *Mean Dawid–Sebastiani score $\overline{DSS}$ and mean ranked probability score $\overline{RPS}$ for the mortality data of Danish and Norwegian women*

	APC	aPC	ApC	APc	apC	aPc	Apc	apc
$\overline{\text{RPS}}$	231.73	134.85	226.14	122.00	120.14	**86.81**	134.50	89.08
$\overline{\text{DSS}}$	11.09	10.34	11.08	10.29	10.27	**9.91**	10.30	9.93

The smallest value for each score is indicated in bold.

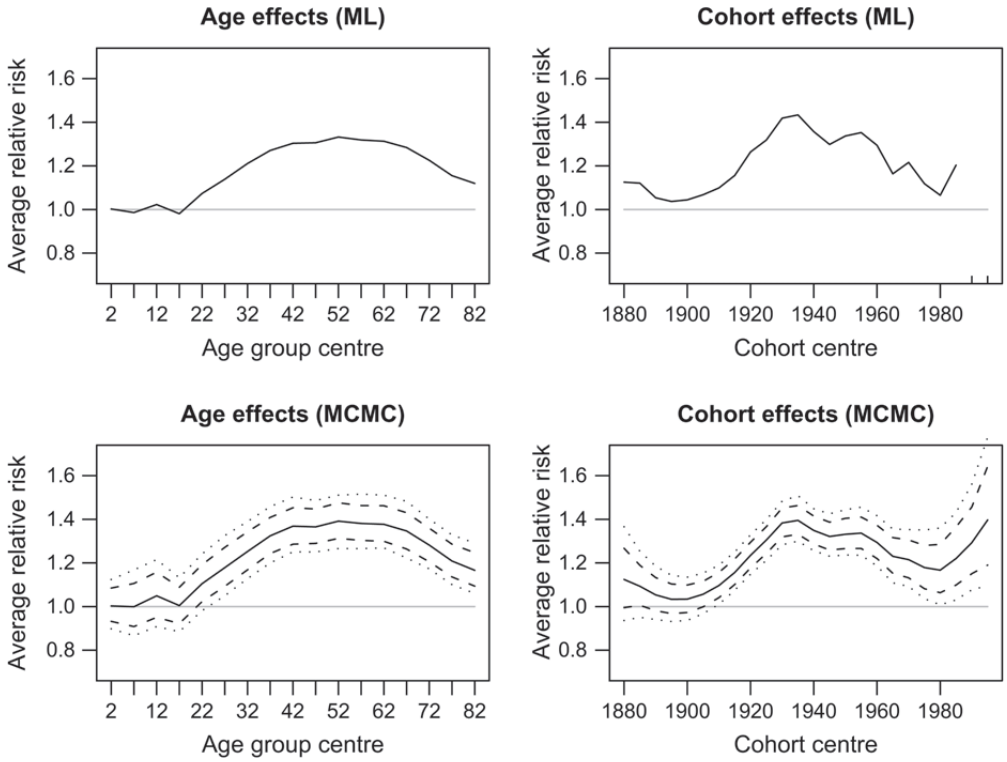

Fig. 1. Average relative risk of death for Danish compared with Norwegian women analyzed by a classical (upper panels) and Bayesian (lower panels) aPc model. Missing ML estimates are indicated by tick marks on the x-axis. For the Bayesian model, the posterior median (solid line) within 95% pointwise (dashed) and simultaneous (dotted) credible bands is shown.

(Niederlaender, 2006). Further details on this analysis can be found in the supplementary material, available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org).

3. Multivariate APC models with unequal time intervals

The approach described above is not applicable when age group and period do not have the same interval lengths, a very common feature of registry data. However, a slightly different definition of the cohorts as proposed in Heuer (1997) can be used to address this problem (see also Knorr-Held and Rainer, 2001). To be more specific, suppose that age is given in M-year intervals, whereas period is given on an annual basis. Cohort indices are then defined by $k = M \times (I - i) + j$. Therefore, a cohort index k can only appear every Mth period. This is illustrated in Table 2 for $M = 10$, which corresponds to the data analyzed in Section 3.2. For example, all cohort parameters with index $k = 1, 11, 21, \ldots, 101$ only appear in the years 1950, 1960, ..., 1990. This particular feature of the APC model with unequal time trends induces additional identifiability problems, as discussed in Section 3.1.

Apart from this additional issue, the multivariate analysis of APC models proceeds as in the case of equal interval lengths. For a Bayesian analysis, we use sum-to-zero constraints on all main effects and do not incorporate any further restrictions. However, only the Bayesian approach based on smoothing priors is now capable to provide sound estimates of relative risk. This is illustrated in the application considered in Section 3.2.

Table 2. *Illustration of the definition of cohort indices for age groups and periods of different interval lengths ($I = 7$, $J = 50$, $K = 110$, $M = 10$)*

	Period (years)													
Age group	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	...	1999
75–84	1	2	3	4	5	6	7	8	9	10	11	12	...	50
65–74	11	12	13	14	15	16	17	18	19	20	21	22	...	60
55–64	21	22	23	24	25	26	27	28	29	30	31	32	...	70
45–54	31	32	33	34	35	36	37	38	39	40	41	42	...	80
35–44	41	42	43	44	45	46	47	48	49	50	51	52	...	90
25–34	51	52	53	54	55	56	57	58	59	60	61	62	...	100
15–24	61	62	63	64	65	66	67	68	69	70	71	72	...	110

3.1 *Identifiability problems of ML estimates*

We first note that in the unequal case, an ML analysis will provide a perfect fit in $2 \times M \times R$ cells. Problems with zero counts will typically increase as the number of observations per cohort parameter decreases with increasing M. The cohort parameters will tend to be unstable if the underlying data are sparse.

A more serious problem arises in the case of unequal time intervals, which is described in length in Holford (2006). We illustrate the problem in the example described in Table 2 where $M = 10$ ($I = 7$, $J = 50$, $K = 110$). Suppose we want to fit the univariate model (2.1) to data available on this resolution. In addition to the unidentifiabilities described earlier, 9 more parameters have to be set to zero, even if there are no problems with zero counts. Indeed, the transformation

$$\varphi_j = \varphi_j + b_{1+(j-1) \bmod 10},$$

$$\psi_k = \psi_k - b_{1+(k-1) \bmod 10}$$

for any real numbers $b_1, \ldots, b_{10}$ subject to $b_1 + \cdots + b_{10} = 0$ will leave the linear predictor η_{ij} unchanged and will maintain the sum-to-zero constraints for period and cohort effects mentioned earlier.

This additional identifiability problem induces artificial cyclical patterns in the period and cohort estimates, which makes the interpretation of ML estimates very difficult. As noted by Holford (2006), the cyclical nature of the additional identifiability problem can give rise to results that can be especially misleading. Cycles that appear in the analysis of data in which the interval widths are unequal must be very carefully analyzed, especially if the periodicity is equal to M. In the multivariate Apc model, these problems further increase with $(M - 1) \times R$ additional period or cohort parameters which have to be set to zero. The remaining ML estimates of period and cohort effects will typically show strong periodicities which makes ML estimates of relative risk impossible to interpret. Note that these problems do not arise when only one set of parameters (period or cohort effects) varies across strata.

3.2 *Application: COPD mortality in England and Wales*

We now analyze annual data on COPD mortality of males in England and Wales, 1950–1999, stratified by $R = 3$ different areas: Greater London, conurbations excluding Greater London, and nonconurbations. Data were obtained by 10-year age bands 15–24, 25–34, ..., 75+, resulting in $I = 7$ age groups, $J = 50$ periods, and $K = 110$ birth cohorts per region (Hansell *and others*, 2003; Hansell, 2004). COPD is a serious lung disease making it difficult to breathe as a consequence of limited airflow. The major risk factor is smoking, which is known to exert mainly long-term effects. A lag period of about 20–30 years between changes in smoking behavior and changes in COPD mortality is suggested (Kazerouni *and others*, 2004). However, smoking trends have not fully explained the higher COPD rates either in UK cities or in more

northern versus more southern areas of England (Reid and Fletcher, 1971; Law and Morris, 1998). Aside of smoking, further risk factors involved in COPD development include air pollution and exposure to dust, gas, or chemical fumes. Air pollution can initiate both long-term (period or cohort) effects and short-term (period) effects (Sunyer, 2001; Dockery and Pope, 1994). Since we are interested in the relation of marked changes in UK air pollution over the last years and changes in COPD mortality, we focus on short-term effects in this application.

Note that an analysis of equal time intervals would either require the aggregation of the yearly data to 10 years resulting in 5 rather than 50 distinct period parameters (and 11 rather than 110 cohort parameters) or the split of the 10-year age groups to 1-year age groups. The first approach severely limits the ability to investigate the timing of period or cohort change points, while the second is expected to introduce only random noise.

Table 3 gives $\overline{\text{DSS}}$ and $\overline{\text{RPS}}$ estimates for all possible models. The apc model is regarded as the best model. However, in the apc model relative risks are not identifiable, compare Section 2.2. Additionally, the use of varying age effects might be questionable because of some specific assumptions in the data generation process. In particular, some data for younger age groups were published only in 20-year age bands between 1950 and 1962. These data were partitioned into 10-year age bands assuming that age breakdowns are the same in all 3 strata (Hansell, 2004). The Apc model is only slightly worse in terms of $\overline{\text{DSS}}$ and $\overline{\text{RPS}}$ and is therefore discussed in the following.

All results are presented relative to nonconurbations as these areas were thought to have had the lowest exposure of factors involved in COPD mortality. Migration is not taken into account, that is, we assumed that people have experienced lifetime air pollution in the region they died. The estimated overdispersion parameters obtained by the ML analysis range from 1.91 to 33.30 for the different models indicating substantial overdispersion. ML and MCMC estimates of average relative risk are shown in Figure 2. In the upper panels, period and cohort estimates obtained from the ML analysis are shown. For both we see extremely large values and artificial cyclical patterns every 10 years. As discussed in Section 3.1, the results are not interpretable.

In contrast to the ML, the Bayesian analysis with smoothing priors provides identifiable and interpretable results as shown in the lower panels of Figure 2. The observed higher average relative risk $\exp(\tilde{\Delta}_j)$ of conurbations compared to Greater London might be due to the fact that conurbations are mainly situated in northern England, where the former predominance of heavy industries resulted in especially high levels of air pollution. The colder climate in the north might also be a reason for the differences (Law and Morris, 1998; Hansell, 2004). There is substantial year-to-year variation in the estimated average relative risk $\exp(\tilde{\Delta}_j)$ from 1950 ($j = 1$) to 1999 ($j = 50$) with higher values in years of known air pollution episodes. For example, the increased average relative risk in 1952 is probably related to the 1952 "Great Smog" in London. Since influenza can exacerbate COPD, the increased average relative risk in 1976 might be caused by the severe influenza epidemic in the same year (Hansell, 2004; Wedzicha, 2004). From the end of the 1970s until 1999, there are almost no changes in average relative risk for conurbations and Greater London.

The average relative risk of cohort effects differs between Greater London and conurbations. For cohorts centers between 1870 and 1880, an increased average relative risk is apparent for Greater London.

Table 3. *Mean Dawid–Sebastiani score $\overline{DSS}$ and mean ranked probability score $\overline{RPS}$ for COPD mortality among males in England and Wales*

	APC	aPC	ApC	APc	apC	aPc	Apc	apc
$\overline{\text{RPS}}$	51.52	39.37	50.42	41.26	30.74	27.45	21.23	**20.13**
$\overline{\text{DSS}}$	7.19	6.82	7.14	7.01	6.56	6.56	6.30	**6.18**

The smallest value for each score is indicated in bold.

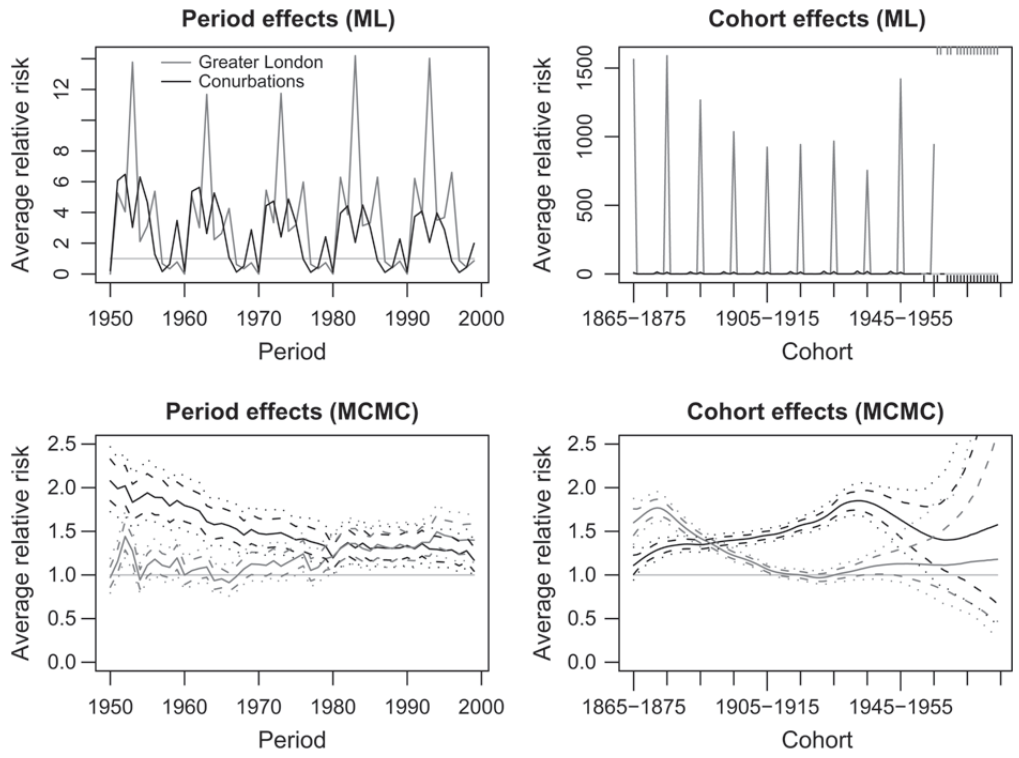

Fig. 2. Average relative risk of death for Greater London and conurbations excluding Greater London compared with nonconurbations analyzed by a classical and Bayesian Apc model. Missing ML estimates are indicated by tick marks on the upper (for Greater London) and lower (for conurbations excluding Greater London) x-axis. For the Bayesian analysis, same quantities as in Figure 1 are shown.

However, with successive cohorts the average risk of Greater London falls from being almost twice as high compared to nonconurbations for the oldest cohorts to being very similar for the youngest cohorts. For conurbations, the average relative risk increases with successive cohorts reaching a maximum for cohorts born around 1940. Possible reasons for the shifted occurrence of pronounced estimates could be different smoking behavior in Greater London and other conurbations.

4. DISCUSSION

In this work, we proposed a Bayesian approach to estimate multivariate APC models. We first illustrated the Bayesian methodology on data of female mortality in Denmark and Norway where age groups and periods are defined on the same time intervals. In this example, the statistical analysis by Poisson regression and the Bayesian approach lead to similar estimates.

In a second application on male COPD mortality in England and Wales where the widths of age and period intervals are unequal, problems using ML became obvious. Especially in the case where both period and cohort effects were allowed to vary across strata, we noticed a cyclical pattern also reported by Holford (2006) for univariate applications on an unequal time grid. The periodicity was equal to the ratio $M = 10$ of the widths of the age group and period intervals making it impossible to disentangle the identifiability problem from what may or may not be a real cyclical pattern. This identifiability problem was resolved through the use of smoothing priors in a Bayesian framework. An alternative approach may be to use splines (Holford, 2006). However, the arbitrariness where to place the knots makes this approach

less attractive in the multivariate case. Penalized splines (Eilers and Marx, 1996) avoid this problem and provide data-driven estimates of smoothing parameters using a mixed-model formulation. However, the number of smoothing parameters in model (2.2) remains a challenge for a larger number of strata.

Somewhat similar in spirit are spatial extensions of Bayesian APC models, which introduce additional spatial effects and possibly space–time interactions to model regional variations of mortality or morbidity rates (Lagazio *and others*, 2003; Schmid and Held, 2004). The spatial effects can be seen as surrogates for factors that influence people living in the same area. The interaction between period and space, or cohort and space, is modeled via suitable prior distributions inducing only temporal (Type II) or spatiotemporal (Type IV) dependence (Knorr-Held, 2000). Our formulation (2.2) can perhaps be seen as a spatial APC model with a Type II interaction prior based on RW2 for both period $\times$ space and cohort $\times$ space but without the main period and cohort effects. In our applications, a meaningful Type IV interaction prior is not possible due to the small number of regions. In contrast, Lagazio *and others* (2003) considered a problem with many regions and used the Type IV prior, which requires a large number of additional sum-to-zero constraints to ensure identifiability. They used RW1 rather than RW2 for the main effects and the associated interaction priors to ensure parameter identifiability. However, the suitability of RW1 for smoothing APC models is not universally accepted as they penalize the first (rather than the second) differences which are not identifiable from the likelihood. Schmid and Held (2004) used both RW1 and RW2 priors but considered only space–period or space–cohort interactions but not both. To summarize, spatial APC models are in fact more complex than our proposed formulation, due to additional adjustments necessary for spatial and spatiotemporal dependence. On the other hand, no emphasis is given on time trends in relative risk or on comparing the results with those obtained by ML.

Although we have analyzed data stratified by geographical region, multivariate APC models can be used in a wider range of applications. For example, different causes of mortality could be analyzed relative to total mortality or male disease rates relative to female disease rates. Of further interest is the analysis of rates stratified by more than one variable. For example, we might be interested to analyze COPD mortality data from men and women in the 3 regions Greater London, conurbations, and nonconurbations. A potentially useful model would allow age effects to be different across gender, while period and cohort effects will be different across regions. Such a nested formulation will detect not only regional differences but also gender-specific variations. We also plan to integrate Bayesian model selection and model averaging in our approach. Using reversible jump MCMC, the appropriate multivariate APC model could be detected for data given in many strata, for example, many small geographical areas. This will allow the analysis to investigate which regions have identical and which have separate effects.

Supplementary Material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

The authors express thanks to Rune Jacobsen and Anna Hansell for access to the data set on women's mortality in Denmark and Norway and COPD mortality in England and Wales, respectively. They also wish to thank Matthias Bopp for helpful discussions and 2 anonymous reviewers for valuable comments on an earlier draft of this manuscript. *Conflict of Interest:* None declared.

Funding

Swiss National Science Foundation (205321-116776, 205320-124429).

REFERENCES

BAKER, A. AND BRAY, I. (2005). Bayesian projections: what are the effects of excluding data from younger age groups? *American Journal of Epidemiology* **162**, 798–805.

BERZUINI, C. AND CLAYTON, D. (1994). Bayesian analysis of survival on multiple time scales. *Statistics in Medicine* **13**, 823–838.

BERZUINI, C., CLAYTON, D. AND BERNARDINELLI, L. (1993). Bayesian inference on the Lexis diagram. *Bulletin of the International Statistical Institute* **50**, 149–164.

BESAG, J., GREEN, P., HIGDON, D. AND MENGERSEN, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3–66.

BRAY, I. (2002). Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **51**, 151–164.

CLAYTON, D. AND SCHIFFLERS, E. (1987). Models for temporal variation in cancer rates. II: age-period-cohort models. *Statistics in Medicine* **6**, 469–481.

CZADO, C., GNEITING, T. AND HELD, L. (2009). Predictive model assessment for count data. *Biometrics* (in press).

DOCKERY, D. W. AND POPE, C. A. (1994). Acute respiratory effects of particulate air pollution. *Annual Review of Public Health* **15**, 107–132.

EILERS, P. H. C. AND MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–102.

FAHRMEIR, L. AND KNORR-HELD, L. (2000). Dynamic and semiparametric models. In: Schimek, M. G. (editor), *Smoothing and Regression: Approaches, Computation and Application*. New York: John Wiley & Sons, pp. 513–544.

FRÜHWIRTH-SCHNATTER, S., FRÜHWIRTH, R., HELD, L. AND RUE, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing* (in press).

FRÜHWIRTH-SCHNATTER, S. AND WAGNER, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika* **93**, 827–841.

GNEITING, T. AND RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**, 359–378.

HANSELL, A., KNORR-HELD, L., BEST, N., SCHMID, V. AND AYLIN, P. (2003). COPD mortality trends 1950–1999 in England & Wales—did the 1956 Clean Air Act make a detectable difference? *Epidemiology* **14**, S55.

HANSELL, A. L. (2004). *The Epidemiology of Chronic Obstructive Pulmonary Disease in the UK: Spatial and Temporal Variations*, [PhD. thesis]. London: Faculty of Medicine of the University of London, Imperial College, St Mary's Campus.

HELWEG-LARSEN, K., KNUDSEN, L. B. AND PETERSSON, B. (1998). Women in Denmark—why do they die so young? Risk factors for premature death. *Scandinavian Journal of Social Welfare* **7**, 266–276.

HEUER, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics* **53**, 161–177.

HOLFORD, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics* **39**, 311–324.

HOLFORD, T. R. (1998). Age-period-cohort analysis. In: Armitage, P. and Colton, T. (editors), *Encyclopedia of Biostatistics*. Chichester: Wiley, pp. 82–99.

HOLFORD, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in Medicine* **25**, 977–993.

JACOBSEN, R., VON EULER, M., OSLER, M., LYNGE, E. AND KEIDING, N. (2004). Women's death in Scandinavia—what makes Denmark different? *European Journal of Epidemiology* **19**, 117–121.

KASS, R. E., CARLIN, B. P., GELMAN, A. AND NEAL, R. M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *American Statistician* **52**, 93–100.

KAZEROUNI, N., ALVERSON, C. J., REDD, S. C., MOTT, J. A. AND MANNINO, D. M. (2004). Sex differences in COPD and lung cancer mortality trends—United States, 1968–1999. *Journal of Womens Health* **13**, 17–23.

KESTELOOT, H. (2001). Queen Margrethe II and mortality in Danish women. *Lancet* **357**, 871–872.

KNORR-HELD, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* **19**, 2555–2567.

KNORR-HELD, L. AND RAINER, E. (2001). Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction. *Biostatistics* **2**, 109–129.

KNORR-HELD, L. AND RUE, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics* **29**, 597–614.

LAGAZIO, C., BIGGERI, A. AND DREASSI, E. (2003). Age-period-cohort models and disease mapping. *Environmetrics* **14**, 475–490.

LAW, M. AND MORRIS, J. (1998). Why is mortality higher in poorer areas and in more northern areas of England and Wales?. *Journal of Epidemiology and Community Health* **52**, 344–352.

MARSHALL, E. C. AND SPIEGELHALTER, D. J. (2003). Approximate cross-validatory predictive checks in disease mapping models. *Statistics in Medicine* **22**, 1649–1660.

NIEDERLAENDER, E. (2006). Causes of death in the EU. Population and social conditions—statistics in focus. Eurostat (European Communities), Catalogue number: KS-NK-06-010-EN-N. Available from: http://epp.eurostat.ec.europa.eu [cited 2009 Feb 12].

PLUMMER, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics* **9**, 523–539.

REID, D. D. AND FLETCHER, C. M. (1971). International studies in chronic respiratory disease. *British Medical Bulletin* **27**, 59–64.

RUE, H. AND HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman & Hall/CRC Press.

SCHMID, V. AND HELD, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics* **60**, 1034–1042.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. R. AND VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B* **64**, 583–639.

SUNYER, J. (2001). Urban air pollution and chronic obstructive pulmonary disease: a review. *European Respiratory Journal* **17**, 1024–1033.

WEDZICHA, J. A. (2004). Role of viruses in exacerbations of chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society* **1**, 115–120.

[*Received August 5, 2008; first revision March 5, 2009; second revision May 5, 2009; third revision July 10, 2009; accepted for publication September 11, 2009*]

Supplementary material to

The analysis of heterogeneous time trends in multivariate age-period-cohort models

Andrea Riebler and Leonhard Held

Biostatistics Unit, Institute of Social and Preventive Medicine,

University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

Tel: +41 44 6344640, FAX: +41 44 6344986, Email: held@ifspm.uzh.ch

Mortality of Danish and Norwegian women

A lively discussion in the press was aroused when Kesteloot (2001) speculated that the low life expectancy of Danish women could be related to the role model effect of Queen Margrethe II, who has been a known cigarette smoker. Here, we re-analyse overall mortality data of Danish and Norwegian women in the years 1960-1999 to search for clues for the low life expectancy of women in Denmark (Jacobsen *and others*, 2004). If the queen's role model effect is indeed linked with the high mortality of Danish women we would expect a pronounced period effect after her ascension to the throne in 1972.

Data are aggregated into 5-year intervals, resulting in $I = 17$ age groups (0-4, 5-9, ..., 80-84), $J = 8$ periods (1960-1964, 1965-1969, ..., 1995-1999) and $K = 24$ cohorts (1875-1884, 1880-1889, 1885-1894, ... 1990-1999) each spanning 10 years whereby two adjacent cohorts overlap by 5 years. In the following uppercases letters for age (A), period (P) or cohort (C) denote a model where the corresponding effect is assumed to be the same for the two strata Denmark and Norway, while

lowercase letters (a, p or c) denote a model where the effect is allowed to differ across strata. We will first present the results of the classical multivariate age-period-cohort analysis using standard Poisson regression and then the results of the Bayesian approach.

Relative risks of death for Danish women compared with Norwegian women obtained through a ML analysis are shown in Figure 1. The (average) relative risk $\exp(\tilde{\Delta}_j)$ in 1960-1964 ($j = 1$), ..., 1995-1999 ($j = 8$) is roughly linear for the ApC, apC and Apc model. However, it increases slowly with time in the ApC and apC model while it decreases in the Apc model. In accordance with Jacobsen *and others* (2004) the APc and Apc model both detected a substantially increased (average) relative risk $\exp(\tilde{\Delta}_k)$ for Danish women born around 1930. In the aPc model the pattern of the average relative risk $\exp(\tilde{\Delta}_k)$ differs slightly as it does not decrease for cohorts born after 1940. Turning to the (average) relative risks $\exp(\tilde{\Delta}_i)$, the aPC, aPc and apC model show a similar pattern, namely an increase of the (average) relative risk for women aged between 40 and 65/70 years.

The estimated overdispersion parameters obtained from the ML analysis are between 4.83 and 64.26 reflecting substantial overdispersion. Adjustments for overdispersion are obviously necessary.

As indicated by small tickmarks on the x-axis of Figure 1 the ML approach did not provide estimates for all periods and cohorts. When computing the average relative risks $\exp(\tilde{\Delta}_i)$ in the case of the apC and aPc model, or $\exp(\tilde{\Delta}_j)$ in the case of the Apc model, the missing values had to be linearly extrapolated to provide results comparable with the Bayesian approach discussed below.

Turning to the Bayesian analysis we compared the models using the mean Dawid-Sabastiani score $\overline{\text{DDS}}$ and the mean ranked probability score $\overline{\text{RPS}}$. For

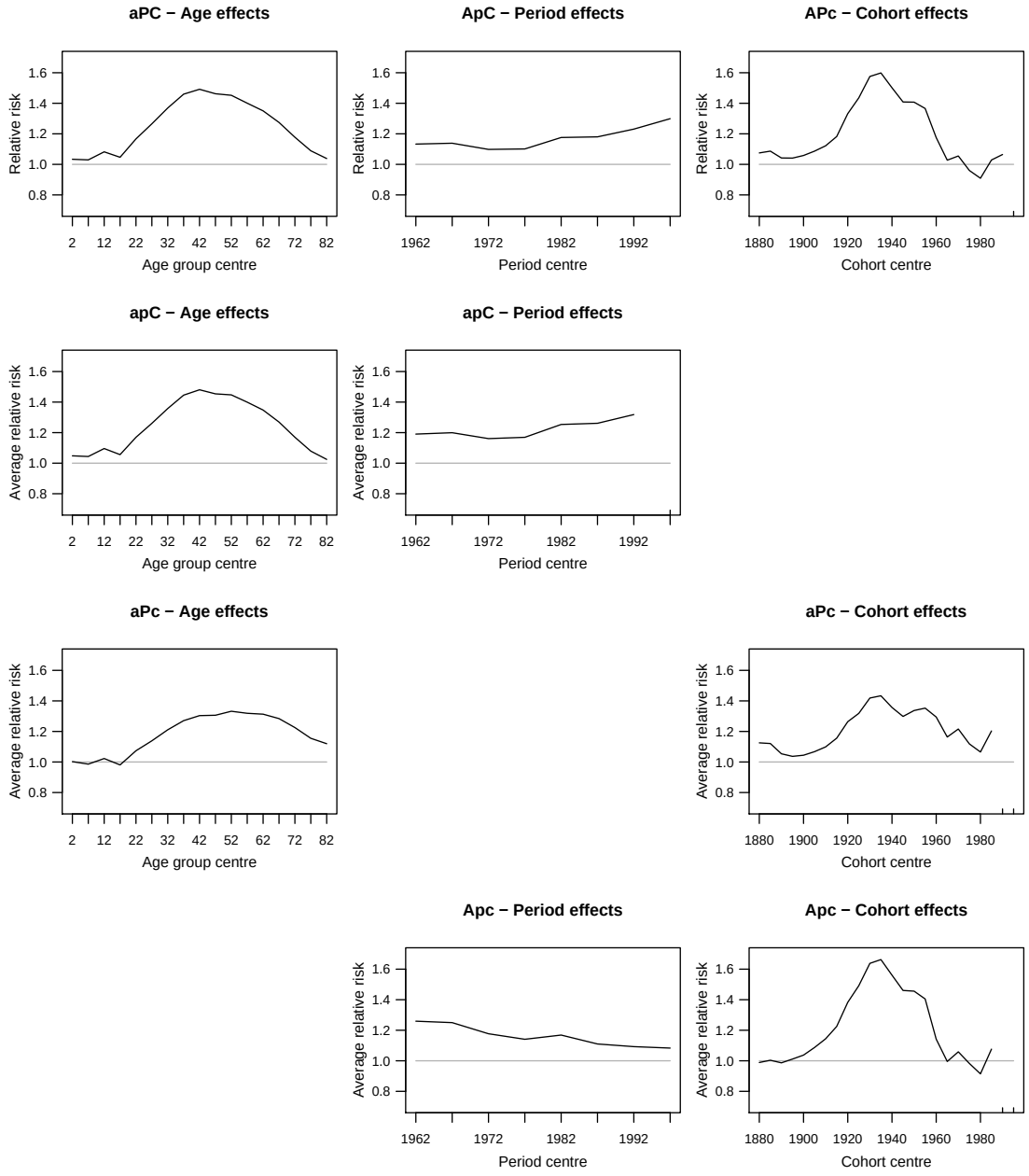

Figure 1: (Average) relative risk of death for Danish compared with Norwegian women based on ML estimates. Missing values are indicated by tickmarks at the corresponding time points on the x-axis.

	APC	aPC	ApC	APc	apC	aPc	Apc	apc
$\overline{\text{RPS}}$	231.73	134.85	226.14	122.00	120.14	**86.81**	134.50	89.08
$\overline{\text{DSS}}$	11.09	10.34	11.08	10.29	10.27	**9.91**	10.30	9.93

Table 1: Mean Dawid-Sabastiani score $\overline{\text{DDS}}$ and mean ranked probability score $\overline{\text{RPS}}$ for the mortality data of Danish and Norwegian women.

both scores smaller values are to be preferred. Table 1 shows these quantities for all possible models.

If allowing one of age, period and cohort to vary, allowing cohort to vary across strata was selected as the best model. When two effects are allowed to vary the model with separate age effects and cohort effects, but joint period effects (model aPc) was assessed to be the best. Quite remarkably, the mean scores $\overline{\text{RPS}}$ and $\overline{\text{DSS}}$ of the aPc model were even smaller than in the most complex model, the apc model, which does not allow to estimate relative risks.

Figure 2 displays posterior median and 95% pointwise and simultaneous credible regions for all models. The 95% simultaneous credible band for the ApC model allows to fit a horizontal line at any relative risk between 1.22 and 1.27. Thus a time-constant relative risk seems to be not unreasonable and this agrees with the only slightly lower $\overline{\text{RPS}}$ and $\overline{\text{DSS}}$ values of the ApC model in comparison to the APC model in Table 1. In none of the models, ApC, apC or Apc we could see evidence for time-changing relative risks in the calendar time scale, which would support the speculations of Kesteloot (2001) regarding the role model effect of Queen Margrethe II of Denmark inducing an increase of mortality in Danish women. In the case of the Apc model a horizontal line fits at any average relative

risk between 1.16 and 1.21.

However, as in the ML analysis there are strong non-linear time trends in the age and cohort time scale. The (average) relative risk is particularly increased for women aged between 40 and 65/70 years as well as for women born after 1925. The difference in age effects might be caused by a different composition of major causes of death in the two countries. The curve progression of $\exp(\tilde{\Delta}_i)$ reminds of the proportion of deaths due to cancer stratified by age groups in the EU (Niederlaender, 2006). Over the years 1955 to 1989 the mortality rates from all cancer diseases over the age from 35 to 64 years were clearly higher (with a ratio of 1.3) in Denmark compared to Norway (Helweg-Larsen *and others*, 1998). Juel *and others* (2000) compared the life expectancy of Danish women aged 35-75 years with Norwegian women in the same age for the years 1991-1993. The difference in life expectancy is 27 months, where all cancer diseases are responsible for 12 months, heart diseases for 6 months and COPD for 3 months. A possible explanation for the striking difference in age effects might therefore be a higher proportion of deaths from cancer in the relevant age groups between 40 and 65/70 years for women in Denmark.

In the case of the (average) relative risk $\exp(\tilde{\Delta}_k)$ the APc and Apc model show decreasing relative risks for women born after 1940. In contrast, the simultaneous credible regions of the aPc model which was assessed best indicate that a time-constant cohort effect might be plausible for women born after 1940. To explain the difference in cohort effects of the Apc model, Jacobsen *and others* (2004) compared the percentage of smokers among Danish and Norwegian women by birth cohort. The overall smoking level of Danish women was higher over all birth cohorts considered. In addition, the cumulative smoking prevalence for Danish women

born 1920-1929 was higher than for the other cohorts. Smoking might therefore be the main causal factor of the observed relative cohort effect around 1930. Since smoking is a risk factor for cancer, heart diseases and COPD, the difference in age-effects might be also related to the much higher smoking prevalence in Denmark, e.g. Danish women are more exposed to risk factors leading to death.

It may also be of interest to investigate the relative risks $\exp(\tilde{\Delta}_{ik})$ in the aPc model. One possibility to visualise those are contour plots, see Figure 3. Here, the mixture of age and cohort effects is apparent. For age groups below the age of 20 there is a horizontal pattern with very low relative risks. The increased relative age effect is not yet apparent and the cohort effect has also not yet started. After the age of 30 there is a horizontal shift visible indicating higher risk for the successive age groups. In addition, a diagonal pattern reflecting the cohort effect for women born after 1925 appears. This effect decreases and after the age of 70 a horizontal pattern reflecting the decreasing relative age effect dominates.

References

Helweg-Larsen, K., Knudsen, L. B. and Petersson, B. (1998). Women in Denmark - why do they die so young? Risk factors for premature death, *Scandinavian Journal of Social Welfare* **7**(4): 266–276.

Jacobsen, R., von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different?, *European Journal of Epidemiology* **19**: 117–121.

Juel, K., Bjerregaard, P. and Madsen, M. (2000). Mortality and life expectancy in

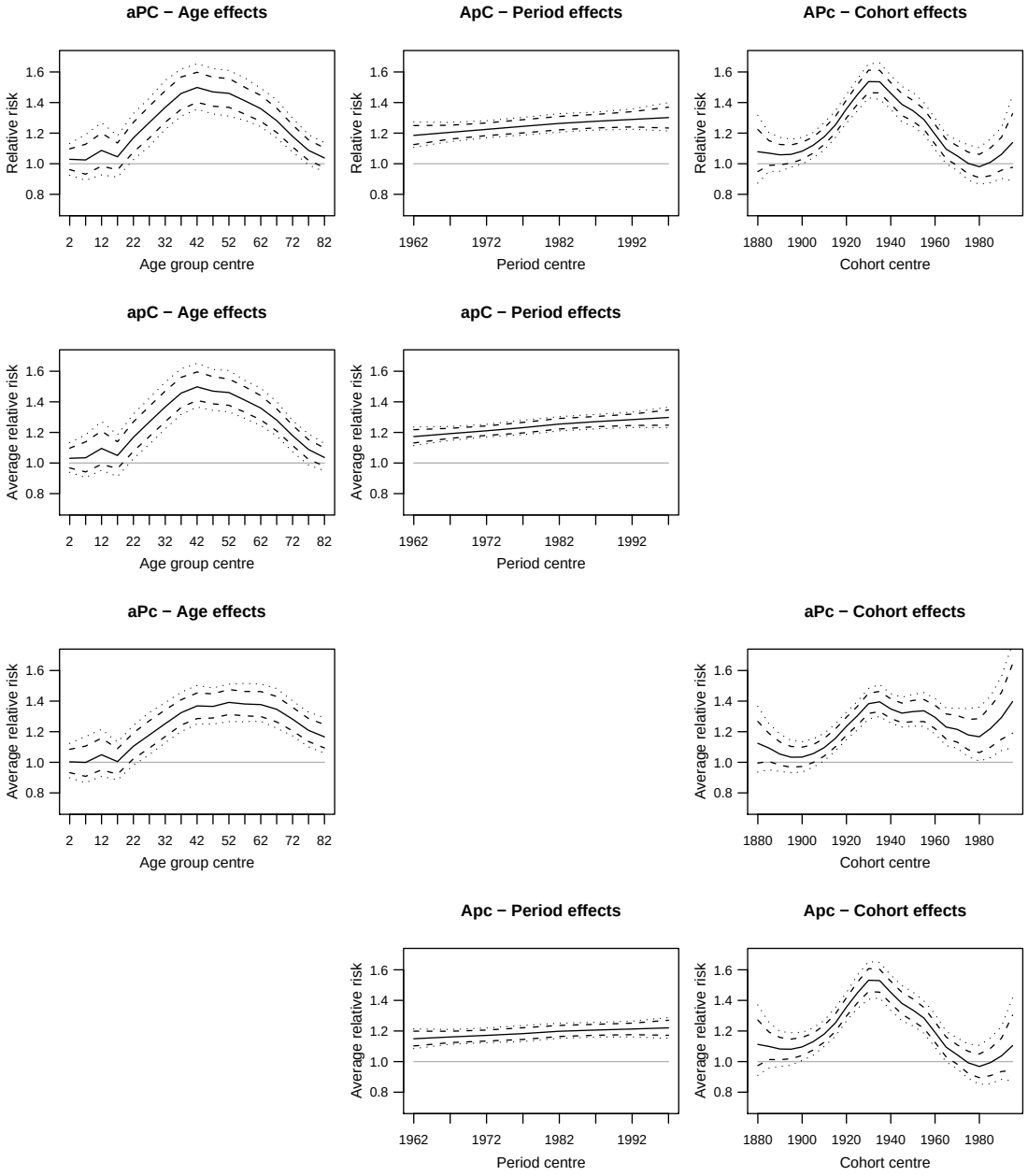

Figure 2: (Average) relative risk of death for Danish compared with Norwegian women analysed by a Bayesian model. Posterior median (solid line) within 95% pointwise (dashed) and simultaneous (dotted) credible bands.

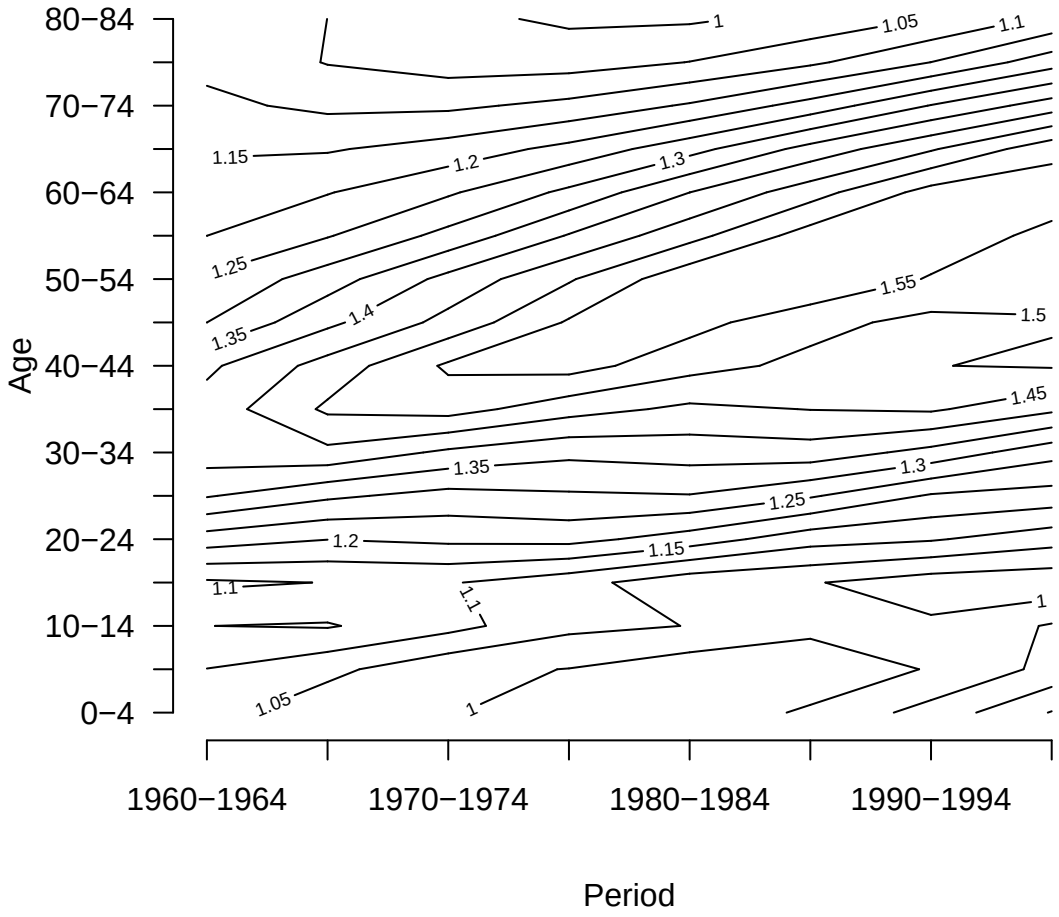

Figure 3: Contour plot of the posterior median relative risk of death for Danish women relative to Norwegian women analysed by the Bayesian aPc model.

Denmark and in other European countries - What is happening to middle-aged Danes?, *European Journal of Public Health* **10**(2): 93–100.

Kesteloot, H. (2001). Queen Margrethe II and mortality in Danish women, *Lancet* **357**: 871–872.

Niederlaender, E. (2006). Causes of death in the EU. Population and social conditions - Statistics in focus. Eurostat (European Communities), Catalogue number: KS-NK-06-010-EN-N. Available from: `http://epp.eurostat.ec.europa.eu` [cited 12.02.2009].

Supplementary material to

The analysis of heterogeneous time trends in multivariate age-period-cohort models

Andrea Riebler and Leonhard Held

Biostatistics Unit, Institute of Social and Preventive Medicine,
University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland
Tel: +41 44 6344640, FAX: +41 44 6344986, Email: `held@ifspm.uzh.ch`

Program manual

This is a short description of the program "`mapc`" for the analysis of multivariate age-period-cohort models. The program was developed under Kubuntu 8.04 (Hardy Heron) on a laptop with Intel(R) Core(TM) 2 Duo T7200 processor 2.0GHz. The program was written in C and uses the GNU Scientific Library (Galassi *and others*, 2009), a numerical library for C and C++, and `GMRFLib` (Rue and Held, 2005, Appendix B), a library in C for fast and exact simulation from Gaussian Markov random fields. Both a precompiled static binary for Linux and the source code of the `mapc` program are provided. The direct use of the binary version, compiled with the provided Makefile, does not require the installation of external libraries.

After downloading or installing the program the components of the model are to be specified in an ini-file. Then start the program by typing

```
./mapc ini-file
```

in the terminal.

Format of the input files

For each stratum r two input files are necessary:

- One file that includes the number of persons at risk n_{ijr} in age group i $(i = 1, \ldots, I)$ and period j $(j = 1, \ldots, J)$.
- One file that includes the number of cases y_{ijr} in age group i $(i = 1, \ldots, I)$ and period j $(j = 1, \ldots, J)$.

The data must be provided in form of a matrix with I rows and J columns.

Structure of the `ini-file`

The `ini-file` specifies all parameters for the algorithm. It is divided in seven sections. Each section starts with a tag written in squared brackets (*[tag]*). The following sections are to be specified:

The *mcmc* section

This section specifies the general settings of the MCMC algorithm. It consists of the following fields:

seed: The seed used for the random number generator.

Default: 1234

burn_in: The number of burn-in iterations.

Default: 20 000

post_burn_in: The number of iterations after the burn-in.

Default: 100 000

thinning: The thinning interval.

Default: 20

The *data* section

This section specifies parameters of the data to be analysed. The following fields need to be specified:

number_of_strata: The number of strata R.

number_of_age_groups: The number of age groups I. (Note: All strata must have the same number of age groups.)

number_of_periods: The number of periods J. (Note: All strata must have the same number of periods.)

periods_per_age_groups: The number of periods per age group, namely the grid-factor M. For example, $M = 1$ for the case in which age group and period have equally spaced intervals. (Note: The data of all strata must have the same intervals.)

counts_stratum_r: The name of the input file which contains the number of cases y_{ijr} for stratum r, where r is an integer with values $r = 1, \ldots, R$. To be more precise, the field *counts_stratum_1* has to be specified for the first stratum, the field *counts_stratum_2* for the second stratum, etc.

population_stratum_r: The name of the input file which contains the number of persons at risk n_{ijr} for stratum r, where r is an integer with values $r = 1, \ldots, R$. For specification compare *counts_stratum_r*.

outputfolder: The name of the sub-directory where the results are stored.

The *random walk* section

order: Order of the random walk (1 or 2) used for age, period and cohort effects.

The *age effects* section

In this section options for the age effects are specified.

separate: A boolean variable indicating whether the age effects should be the same or should vary across strata. Strings starting with "y", "Y", "t", "T" or "1" can be used to specify true values (return 1), strings starting with "n", "N", "f", "F", "0" represent false values (return 0).

initial : Starting value for the precision.

Default: 10.0

parameter_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameter_b: Rate parameter for the gamma prior of the precision.

Default: 0.000 05

The *period effects* section

In this section options for the period effects are specified.

separate: A boolean variable indicating whether the period effects should be the same or should vary across strata. For details on specification see also *age effects* section.

initial : Starting value for the precision.

Default: 200.0

parameter_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameter_b: Rate parameter for the gamma prior of the precision.

Default: 0.000 05

The *cohort effects* section

In this section options for the cohort effects are specified.

separate: A boolean variable indicating whether the cohort effects should be the same or should vary across strata. For details on specification see also *age effects* section.

initial : Starting value for the precision.

Default: 200.0

parameter_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameter_b: Rate parameter for the gamma prior of the precision.

Default: 0.000 05

The *overdispersion* section

In this section options for the overdispersion are specified.

initial : Starting value for the precision.

Default: 100.0

parameter_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameter_b: Rate parameter for the gamma prior of the precision.

Default: 0.005

Output files

The MCMC algorithm samples from the posterior distributions and stores the generated samples of length $N = \textit{post\_burn\_in}/\textit{thinning}$ in the *outputfolder*. Each output file has N rows. The following output files are generated:

- `prec_overdis_1.txt`: Contains the samples of the precision of the overdispersion.
- `age_1.txt`: Contains the samples of the age effects in I columns, where each column refers to a different age effect.
- `period_1.txt`: Contains the samples of the period effects in J columns.

- `cohort_1.txt`: Contains the samples of the cohort effects in $K = M \times (I - 1) + J$ columns.
- `prec_age_1.txt`: Contains the samples of the precision of the age effects.
- `prec_period_1.txt`: Contains the samples of the precision of the period effects.
- `prec_cohort_1.txt`: Contains the samples of the precision of the cohort effects.

For the case in which the option *separate* in the age, period or cohort effects section was set to true, there are also the corresponding precision and effect files for strata $r = 2, \ldots, R$. (Note: There is no separate precision for the overdispersion.)

For each stratum r, there are the output files:

- `mu_r.txt`: Contains the samples of the intercept of stratum r.
- `eta_r.txt`: Contains the samples of the linear predictor η_{ijr} (including the offset $\log(n_{ijr})$) in $I \times J$ columns:

$$\eta_{11r} \quad \eta_{12r} \quad \ldots \quad \eta_{1Jr} \quad \eta_{21r} \quad \eta_{22r} \quad \ldots \quad \eta_{2Jr} \quad \eta_{31r} \quad \ldots \quad \eta_{IJr}$$

- `z_r.txt`: Contains the samples of the overdispersion z_{ijr} of stratum r in $I \times J$ columns, see also `eta_r.txt`.
- `yrep_r.txt`: Contains the samples of the replicated data points y_{ijr}^{rep} of stratum r used to calculate the mean ranked probability score ($\overline{\text{RPS}}$) and the mean Dawid-Sebastiani score ($\overline{\text{DSS}}$). The file has $I \times J$ columns, see also `eta_r.txt`.
- `expected_r.txt`: contains the samples of the expected counts of stratum r in $I \times J$ columns, see also `eta_r.txt`.

R-functions for model choice

To calculate the mean ranked probability score and the mean Dawid-Sebastiani score, two functions implemented in R (R Development Core Team, 2009) are provided in the file `scoring_rules.R`. As input both functions need the path to the *outputfolder* and a vector with the paths to the data files in which the numbers of cases for each stratum are stored.

Example: Mortality of Danish and Norwegian women (aPc)

In the directory `dk_n` the data files to re-analyse the overall mortality of Danish and Norwegian women are provided (Jacobsen *and others*, 2004). According to our model choice criteria, the model with separate age and cohort effects, but joint period effects (aPc) was classified as the best model.
The corresponding `ini-file`, called `aPc_dk_n_rw2.ini`, is:

```
[mcmc]
seed = 1234
burn_in = 20000
post_burn_in = 100000
thinning = 20

[data]
number_of_strata = 2
number_of_age_groups = 17
number_of_periods = 8
periods_per_age_groups = 1

counts_stratum_1 = ./dk_n/dk_1960_counts.txt
population_stratum_1 = ./dk_n/dk_1960_pop.txt
counts_stratum_2 = ./dk_n/n_1960_counts.txt
population_stratum_2 = ./dk_n/n_1960_pop.txt

outputfolder = ./aPc/

[random walk]
order = 2

[age effects]
separate = yes
initial = 10.0
parameter_a = 1.0
parameter_b = 0.00005

[period effects]
separate = no
initial = 200.0
parameter_a = 1.0
parameter_b = 0.00005

[cohort effects]
separate = yes
```

```
37 initial = 200.0
38 parameter_a = 1.0
39 parameter_b = 0.00005
40
41 [overdispersion]
42 initial = 100.0
43 parameter_a = 1.0
44 parameter_b = 0.005
```

The first section of the `ini-file` specifies general settings for the MCMC algorithm. Here, a burn-in of 20 000 iterations (line 3), followed by 100 000 post-burn-in iterations (line 4) is defined. The ***thinning*** variable (line 5) specifies that the samples of every 20th iteration should be stored.

The number of strata, age groups and periods, as well as the grid factor, e. g. the number of periods per age group are specified in the second section (lines 8-11). Lines 13-16 specify the name of the files where the data are stored. The directory in which the samples will be stored is defined in line 18.

In the third section the order of the random walk is set. Here, a second order random walk is defined (line 21) for age, period and cohort effects.

The following three sections specify the settings of the age, period and cohort effects. The shape and rate parameter of the gamma prior for the precisions are set to 1.0 and 0.000 05 for all effects. Initial values for the precisions are set using the variable ***initial***. The type of model, here an aPc model, is specified using the variable ***separate*** of each section. In the age effects and cohort effects section this variable is set to "yes", while in the period effects section it is set to "no".

In the last section the initial value and parameters of the gamma prior for the precision of the overdispersion are specified.

To run the program, type `./mapc aPc_dk_n_rw2.ini` in the terminal. On a laptop with Intel(R) Core(TM) 2 Duo T7200 processor 2.0GHz the analysis takes about 6 minutes. For calculating the mean ranked probability score and the mean Dawid-Sebastiani score type in R:

```
> source("./scoring_rules.r")
> dss(results_path="./aPc/",
+     data_path= c("./dk_n/dk_1960_counts.txt",
+                  "./dk_n/n_1960_counts.txt"))
```

[1] 9.908019

```
> rps(results_path="./aPc/",
+     data_path= c("./dk_n/dk_1960_counts.txt",
+                  "./dk_n/n_1960_counts.txt"))
```

[1] 86.81348

References

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M. and Rossi, F. (2009). *GNU Scientific Library Reference Manual - Third Edition*, Network Theory Ltd. ISBN 0954612078.

Jacobsen, R., von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different?, *European Journal of Epidemiology* **19**: 117–121.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. Available from: `http://www.R-project.org`.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC Press, London.

PAPER II

A conditional approach for inference in multivariate age-period-cohort models

Leonhard Held & Andrea Riebler

Paper accepted for publication in *Statistical Methods in Medical Research*.

A conditional approach for inference in multivariate age-period-cohort models

Leonhard Held[1] and Andrea Riebler

Biostatistics Unit, Institute of Social and Preventive Medicine,

University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

Tel: +41 44 6344640, FAX: +41 44 6344986, Email: held@ifspm.uzh.ch

Age-period-cohort (APC) models are used to analyse data from disease registers given by age and time. When data are stratified by one further variable, for example geographical location, multivariate APC (MAPC) models can be applied to identify and estimate heterogeneous time trends across the different strata. In such models, outcomes share a set of parameters, typically the age effects, while the remaining parameters may differ across strata. In this paper, we propose a conditional approach for inference to directly model relative time trends. We show that in certain situations the conditional approach can handle unmeasured confounding so that relative risks might be estimated with higher precision. Furthermore, we propose an extension for data with more stratification levels. Maximum likelihood estimation is performed using software for multinomial logistic regression. The usage of smoothing splines is suggested to stabilise estimates of relative time trends, if necessary. We apply the methodology to chronic obstructive pulmonary disease mortality data in England & Wales, stratified by three different areas and gender.

Keywords: Conditional likelihood; Multinomial logistic regression; Multivariate age-period-cohort model; Overdispersion; Relative risk.

[1]To whom correspondence should be addressed.

1 Introduction

Most developed countries collect information on morbidity and mortality rates by recording new cases by diagnosis, gender, age and date of diagnosis. To detect temporal patterns these epidemiological data can be analysed by age-period-cohort (APC) models (Holford, 1983; Clayton and Schifflers, 1987; Holford, 1998). The starting point for APC analyses is the assumption that the outcome under consideration is a result of separate contributions of age, calendar period and birth cohort effects. Age effects seek to explain the changes in rates across different age groups. Factors likely to affect all people at a particular date, regardless of their age, are called period effects, e.g. air pollution or medical advances. In contrast, cohort effects tend to be factors common in people born at a particular period in time, e.g. smoking habits.
It is important to be aware of the well-known non-identifiability problem in APC models (Holford, 1998). Since, the date of diagnosis is the sum of the date of birth and age at diagnosis it is not possible to disentangle the three time trends. There are infinitely many linear transformations that all lead to the same estimated incidence or mortality rates. For this reason, non-linear trends, e.g. changes in slope, are interpretable but linear trends on a time scale are not. To identify the model parameters a further constraint is required (Holford, 1983, 1991; Robertson *et al.*, 1999). The most direct approach is to remove one time effects group, i.e. set all its effects to zero, so that not all three parameter groups are simultaneously in the model. Another possibility is to equate, for example, two period effects. Unfortunately, there is often no sound basis for equating a specific pair of effects. It might be equally appropriate to equate a different pair. However, depending on the chosen pair the resulting parameter estimates might vary considerably (Holford, 1991). For a contrast and comparison of more elaborate approaches to derive the necessary extra linear constraint we refer to Robertson and Boyle (1998), see also Fu (2000) and Yang *et al.* (2008).
Frequently, data are additionally stratified, for example by geographical region. Thus a vector of observations is given for each age group at each period index. For such data multivariate APC (MAPC) models have been considered (Hansell *et al.*, 2003; Hansell, 2004; Jacobsen *et al.*,

2004; Riebler and Held, 2010). A joint analysis may borrow strength from sharing a set of parameters, typically the age effects, while estimating region-specific period and/or cohort effects. Riebler and Held (2010) showed that differences of stratum-specific parameters are identifiable and can be interpreted as log relative risks. They proposed a Bayesian hierarchical model based on a Poisson likelihood for inference in MAPC models.

In this paper, we propose an alternative approach to fit MAPC models, in which we condition on the sum of mortality or morbidity counts over the different strata. Conditioning is a useful technique to remove nuisance parameters from a likelihood, see McCullagh and Nelder (1989, Chapter 7) and Pawitan (2001, Chapter 10). For example, a conditional likelihood can be constructed both for the rate ratio and the odds ratio (Clayton and Hills, 1993). Similarly, Kelsall and Diggle (1998) have proposed a binary regression approach to estimate relative risk in spatial case-control studies based on conditioning. In the case of MAPC models, we shall see that the conditional formulation allows us to model the parameters of interest, the log relative risks, directly.

A further attractive aspect of the conditional approach is that it may reduce the amount of overdispersion typically encountered in registry data. When modelling data of comparable strata, e.g. the same disease in neighbouring regions, it seems plausible that overdispersion is at least partially caused by unobserved explanatory variables acting on all strata simultaneously. In this case more precise relative risks will be obtained through conditioning, for which we will give an intuitive explanation.

The conditional approach leads to a multinomial logistic regression model with stratum-specific offsets. Maximum likelihood estimation and adjustments for overdispersion are straightforward to implement. The optional usage of smoothing splines is suggested to stabilise the cohort-specific relative risk estimates. Smoothing splines are mandatory in the case of unequal age group and period intervals with stratum-specific period and cohort effects to avoid artificial periodicities in the estimated relative risks (Holford, 2006; Riebler and Held, 2010). Model choice can be based on Akaike's information criterion (AIC) or a suitable extension in the case

of overdispersion. Furthermore, the approach can be extended to data stratified by up to three additional stratification variables. The conditional approach is directly applicable and results in separate multinomial logistic regression models to estimate relative risk parameters associated with the different stratification variables.

The paper is organised as follows. Section 2 introduces yearly data on chronic obstructive pulmonary disease (COPD) mortality in England & Wales, 1950-1999, for 7 age groups (Hansell *et al.*, 2003; Hansell, 2004). The dataset is stratified by two additional variables, namely gender and geographical regions and will be used throughout the paper to motivate the use of the conditional approach. In Section 3 we review MAPC models. In Section 4 we describe the new conditional approach to infer relative risk parameters in MAPC models. Section 5 describes the extended model formulation for data with more than one additional stratification level and the conditional approach for inference using separate conditional analyses. In Section 6 we present a detailed analysis of the COPD dataset introduced in Section 2. First, we re-analyse data among males only. In this case study overdispersion is considerably reduced using the conditional approach. As a consequence relative risks can be estimated with higher precision than using the corresponding unconditional formulations. To allow for a joint analysis of males and females, we then consider an extended MAPC model with gender-specific age effects and region-specific period and cohort effects.

2 COPD mortality rates in England & Wales

Chronic obstructive pulmonary disease (COPD) is a serious lung disease making it difficult to breathe. The COPD mortality data we use are provided from 1950-1999 and stratified by 3 regions (Greater London, conurbations excluding Greater London and rural areas) and by gender (Hansell *et al.*, 2003; Hansell, 2004). Age groups are given in ten-year bands: 15-24, 25-34, ldots, 75+, resulting in 7 age groups and 50 periods. Data on males were also analysed in Riebler and Held (2010), who provide further background and references on epidemiological aspects of the disease.

3 Multivariate APC models

We describe our model formulation in terms of the COPD dataset, but of course it applies more generally. To begin, let n_{ijr} denote the number of persons at risk in age group i ($i = 1, \ldots, I$), period j ($j = 1, \ldots, J$) and stratum r ($r = 1, \ldots, R$); here $I = 7$, $J = 50$ and $R = 3$ ($r = 1$: Greater London, $r = 2$ conurbations, $r = 3$ rural areas). We follow Riebler and Held (2010) and assume that the number of cases y_{ijr} in age group i during period j in stratum r has a Poisson distribution with rate $n_{ijr} \cdot \lambda_{ijr}$, and that all observations y_{ijr} are independent, given the unknown relative risk parameters λ_{ijr}. In the most general formulation the linear predictor $\eta_{ijr} = \log\{\lambda_{ijr}\}$ results as the sum of a stratum-specific intercept μ_r and stratum-specific age, period and cohort effects θ_{ir}, φ_{jr} and ψ_{kr}, respectively. Thus

$$\eta_{ijr} = \mu_r + \theta_{ir} + \varphi_{jr} + \psi_{kr}. \tag{3.1}$$

Here, we refer to a particular birth cohort by the index $k = 1, \ldots, K$ which is a linear function of age group index i and period index j. In addition k depends on the ratio M of the widths of the age group and period intervals, so that $k = M \times (I - i) + j$ (Knorr-Held and Rainer, 2001). Note that in our application the age groups are given in 10-years intervals while the periods are provided yearly, so that $M = 10$ and $K = 10 \times (7 - 1) + 50 = 110$. For identifiability of the stratum-specific intercepts, additional sum-to-zero constraints have to be imposed on the time effects, so that we set $\sum_i \theta_{ir} = \sum_j \varphi_{jr} = \sum_k \psi_{kr} = 0$ for all $r = 1, \ldots, R$. However, note that further identifiability problems remain due to the linear dependence between age, period and cohort effects in each stratum.

Simpler model formulations result by allowing for effects that are identical across strata. Throughout the paper we will use lower case letters a, p or c to denote effects which are allowed to vary across strata while upper case letters A, P, or C to denote identical effects. For example, let us consider the Apc model in the following. Here, the age effects are assumed to be identical across the 3 regions of the COPD dataset, while the period and cohort effects are different. The

linear predictor is

$$\eta_{ijr} = \mu_r + \theta_i + \varphi_{jr} + \psi_{kr}. \tag{3.2}$$

It is obvious how to modify this model in the case of unequal period but equal cohort effects (ApC) or vice versa (APc). Similarly, age and cohort effects may vary across the regions but the period effects may be fixed, for example (aPc). However, keep in mind that one further identifiability problem still remains due to the linear dependence between age, period and cohort effects.

Consider now the difference $\Delta_j^{(r)} = \varphi_{j,r} - \varphi_{j,R}$ of the j-th period effect in stratum r ($r = 1, \ldots, R-1$) and the j-th period effect in stratum R. To keep notation simple, we always use the R-th stratum as reference stratum but any other stratum could be used as reference, of course. Riebler and Held (2010) show that $\Delta_j^{(r)}$ is identifiable as long as the strata share the same age effects. Differences $\Delta_k^{(r)} = \psi_{k,r} - \psi_{k,R}$ of two cohort effects are also identifiable.

Let $\Delta_\mu^{(r)} = \mu_r - \mu_R$. In the case where only the period effects are chosen to differ across strata, the adjusted difference $\tilde{\Delta}_j^{(r)} = \Delta_\mu^{(r)} + \Delta_j^{(r)}$ can then be interpreted as the *log relative risk* in period j and stratum r, relative to stratum R. Similarly, if the cohort effects are chosen to differ across strata, $\tilde{\Delta}_k^{(r)} = \Delta_\mu^{(r)} + \Delta_k^{(r)}$ is the log relative risk of cohort k in stratum r, relative to stratum R. If both period and cohort effects are allowed to vary across strata, the log relative risk $\tilde{\Delta}_{jk}^{(r)} = \Delta_\mu^{(r)} + \Delta_j^{(r)} + \Delta_k^{(r)}$ depends both on period j and cohort k. In this case $\tilde{\Delta}_j^{(r)}$ and $\tilde{\Delta}_k^{(r)}$ can be interpreted as *average log relative risk*:

$$\tilde{\Delta}_j^{(r)} = \frac{1}{K} \sum_k \tilde{\Delta}_{jk}^{(r)} \quad \text{and} \quad \tilde{\Delta}_k^{(r)} = \frac{1}{J} \sum_j \tilde{\Delta}_{jk}^{(r)}$$

due to $\sum_k \Delta_k^{(r)} = 0$ and $\sum_j \Delta_j^{(r)} = 0$, respectively. For example, $\tilde{\Delta}_j^{(r)}$ is now the log relative risk in period j, averaged over all cohorts $k = 1, \ldots, K$. Note that we average over all cohorts, not just over those for which data have been observed at time j. This should be kept in mind when interpreting $\tilde{\Delta}_j^{(r)}$. In principle one could average only over the cohorts observed in period j, but then the cohort effects would not cancel. Similarly, $\tilde{\Delta}_k^{(r)}$ is the log relative risk in cohort

k, averaged over all periods $j = 1, \ldots, J$.

Therefore, $\exp(\tilde{\Delta}_j^{(r)})$ is the geometrically averaged relative risk in period j and likewise $\exp(\tilde{\Delta}_k^{(r)})$ in cohort k. For simplicity we will use the term average relative risk (rather than geometrically average relative risk) in the rest of the paper.

Maximum likelihood (ML) estimation in MAPC models can be performed with standard software for Poisson regression. However, interpretation of the results is far from trivial. This is due to the fact that the design matrix for the parameters is singular because of the non-identifiabilities mentioned above. Statistical software will then delete columns in this design matrix until regularity is achieved. Further singularities may arise if there are cell counts with zero cases. Often a considerable number of cohort parameter estimates in a MAPC model will be missing, especially if the data are given in unequal intervals. This is problematic, because the sum-to-zero constraints apply to all parameters of a cohort block, so that it is impossible to apply the constraint if some of them are missing. The interpretation of $\tilde{\Delta}_j^{(r)}$ and $\tilde{\Delta}_k^{(r)}$ as average log relative risk is then difficult. Linear extrapolation of the missing effects might help to recover the relative risk parameters, but it is ad-hoc. Spline smoothing has been suggested (Heuer, 1997) but the linear dependence of age, period and cohort parameters still remains an issue. For example, the computation of (pointwise) confidence bands for the relative risk parameters is very cumbersome.

These problems with maximum likelihood estimates have led Riebler and Held (2010) to propose a Bayesian smoothing approach. Similarly to Berzuini and Clayton (1994) and Besag *et al.* (1995) for the univariate APC model, they used second-order random walk (RW2) smoothing priors independently for the age effects $\boldsymbol{\theta}$, all period effects $\boldsymbol{\varphi}_r$ and all cohort effects $\boldsymbol{\psi}_r$ in (3.2). This is a natural choice as it penalises the second differences, i.e. changes in trend, which are identifiable in the APC model. For example, the prior density on the period effects $\boldsymbol{\varphi}_r$ in stratum r is

$$p(\boldsymbol{\varphi}_r|\kappa_r) \propto \exp\left(-\frac{\kappa_r}{2}\sum_{j=3}^{J}(\varphi_{j,r} - 2\varphi_{j-1,r} + \varphi_{j-2,r})^2\right),$$

where κ_r is the scale parameter that determines the amount of smoothing. Large values of κ_r correspond to a high degree of smoothing, while small values correspond to a small amount of smoothing. The prior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\psi_r}$ take similar forms with different scale parameters. All scale parameters were treated as unknown and highly dispersed but proper gamma prior distributions were assigned.

Estimation of the relative risk parameters is then straightforward, since the smoothing priors make latent parameters still identifiable, even if no data are observed for some values. The only identifiability problem remaining is due to the linear dependence of age, period and cohort parameters. However, the Bayesian approach allows to have unidentifiable posterior quantities, as long as the quantities of interest (the relative risk parameters) are identifiable. This is discussed in detail in Gelfand and Sahu (1999). They show that an embedded lower dimensional parameter vector (in our case the relative risk parameters) may well have a proper posterior, even if the posterior distribution of all parameters in the model is improper. In practice, one can place any proper prior (including deterministic constraints) on nonestimable functions of the parameters, but the posterior of the relative risk parameters will remain the same. See Roberts *et al.* (1995) for additional discussion.

In real life applications, the Poisson assumption may be too restrictive and it will often be necessary to adjust for overdispersion. In a frequentist framework this is easily done with a quasi-likelihood approach (Zheng *et al.*, 1996; Holford, 2006). Here an overdispersion parameter ϕ_{UC}, computed from the regression output, can be used to inflate the variance of the parameter estimates (Breslow, 1984). A full likelihood-based alternative is negative binomial regression (Hilbe, 2007). Within the Bayesian approach, the usual way to adjust for overdispersion is to introduce additional independent random effects in the linear predictor, for example mean-zero normal variables with unknown variance (Besag *et al.*, 1995).

4 A conditional approach

A standard result from probability theory states that, if X and Y are independent Poisson random variables with mean α and β, say, then the conditional distribution of X, given $Z = X + Y = z$ is binomial with denominator z and success probability $\alpha/(\alpha + \beta)$, see for example Casella and Berger (1990, Chapter 4). A generalisation of this result states that a vector of n independent Poisson random variables $X_1, \ldots, X_n$ with mean $\lambda_1, \ldots, \lambda_n$ is, after conditioning on the sum $X_\bullet = X_1 + \ldots + X_n = x_\bullet$, multinomial distributed with denominator $x_\bullet$ and individual success probabilities $\lambda_i/(\sum_{i=1}^n \lambda_i)$. This result can be applied to the original MAPC model in a way such that only the relative risk parameters enter the conditional distribution (e.g. Agresti, 2002, Section 8.6.7). A multinomial logistic regression model can then be used to infer the relative risk parameters.

For example, consider a multivariate MAPC model with $R = 2$ strata and joint age effects but separate period and cohort effects (Apc). Let $y_{ij\bullet} = y_{ij1} + y_{ij2}$. Using (3.2) it follows that $y_{ij1}|y_{ij\bullet}$ is binomial with denominator $y_{ij\bullet}$ and success probability

$$\begin{aligned} \pi_{ij1} &= \frac{n_{ij1}\lambda_{ij1}}{n_{ij1}\lambda_{ij1} + n_{ij2}\lambda_{ij2}} \\ &= \text{expit}\left(\log\left(\frac{n_{ij1}}{n_{ij2}}\right) + \Delta_\mu + \Delta_j + \Delta_k\right), \end{aligned} \tag{4.1}$$

here $\text{expit}(x) = 1/[1+\exp(-x)]$ is the inverse logit function. Note that through conditioning, the original intercepts, period and cohort parameters are replaced by the differences $\Delta_\mu = \mu_1 - \mu_2$, $\Delta_j = \varphi_{j1} - \varphi_{j2}$ and $\Delta_k = \psi_{k1} - \psi_{k2}$. Age effects cancel in this conditional formulation. Equation (4.1) is of the form of a logistic regression model with offset $\log(n_{ij1}/n_{ij2})$.

A conditional likelihood argument can be used to motivate the usage of this logistic regression model rather than the original Poisson regression model. Indeed, following the arguments in Pawitan (2001, Section 10.3) or Clayton and Hills (1993, Section 13), we first transform the data y_{ij1}, y_{ij2} to $y_{ij1}, y_{ij\bullet} = y_{ij1} + y_{ij2}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$ without any loss of information. The probability function of y_{ij1} and $y_{ij\bullet}$ can be factorised into the conditional probability function

of $y_{ij1}|y_{ij\bullet}$, given above, and the marginal probability function of $y_{ij\bullet}$:

$$f(y_{ij1}, y_{ij\bullet}) = f(y_{ij1}|y_{ij\bullet})f(y_{ij\bullet}).$$

An informal argument says that the total number of cases $y_{ij\bullet}$ tells us nothing about the relative risk parameters, hence it should be sufficient to use the conditional distribution alone for inference about the relative risk parameters.

In our example, we even observe the ideal case of orthogonal parameters (see Pawitan, 2001, Section 10.2), so the conditional likelihood carries exactly the same information about the relative risk parameters as the joint likelihood. To see this, let $\boldsymbol{\lambda}_1$ denote all λ_{ij1}'s and likewise $\boldsymbol{\lambda}_2$ all λ_{ij2}'s. Because of conditional independence, the likelihood function implied by the MAPC model introduced in Section 3 can be written as

$$L(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \prod_{i,j} L(\lambda_{ij1}, \lambda_{ij2})$$

with

$$L(\lambda_{ij1}, \lambda_{ij2}) = (n_{ij1}\lambda_{ij1})^{y_{ij1}}(n_{ij2}\lambda_{ij2})^{y_{ij2}} \exp(-(n_{ij1}\lambda_{ij1} + n_{ij2}\lambda_{ij2})). \tag{4.2}$$

Our parameter of interest is the relative risk $\alpha_{ij} = \lambda_{ij1}/\lambda_{ij2}$, and we consider $\beta_{ij} = n_{ij1}\lambda_{ij1} + n_{ij2}\lambda_{ij2}$ as nuisance parameter. Multiplying (4.2) with $1 = \beta_{ij}^{-y_{ij\bullet}}\beta_{ij}^{y_{ij\bullet}}$ we easily see that the likelihood contribution $L(\lambda_{ij1}, \lambda_{ij2})$, expressed in terms of α_{ij} and β_{ij}, is

$$L(\alpha_{ij}, \beta_{ij}) = \left(\frac{\frac{n_{ij1}}{n_{ij2}}\alpha_{ij}}{1 + \frac{n_{ij1}}{n_{ij2}}\alpha_{ij}}\right)^{y_{ij1}} \left(\frac{1}{1 + \frac{n_{ij1}}{n_{ij2}}\alpha_{ij}}\right)^{y_{ij2}} \beta_{ij}^{y_{ij\bullet}} \exp(-\beta_{ij}),$$

so $L(\alpha_{ij}, \beta_{ij}) = L_1(\alpha_{ij})L_2(\beta_{ij})$ with

$$L_1(\alpha_{ij}) = \left(\frac{\frac{n_{ij1}}{n_{ij2}}\alpha_{ij}}{1 + \frac{n_{ij1}}{n_{ij2}}\alpha_{ij}}\right)^{y_{ij1}} \left(\frac{1}{1 + \frac{n_{ij1}}{n_{ij2}}\alpha_{ij}}\right)^{y_{ij2}}$$

$$\text{and } L_2(\beta_{ij}) = \beta_{ij}^{y_{ij\bullet}} \exp(-\beta_{ij})$$

and therefore $L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = L_1(\boldsymbol{\alpha})L_2(\boldsymbol{\beta})$ with

$$L_1(\boldsymbol{\alpha}) = \prod_{i,j} L_1(\alpha_{ij})$$
$$\text{and } L_2(\boldsymbol{\beta}) = \prod_{i,j} L_2(\beta_{ij}).$$

This means that the likelihood implied by the MAPC model can be rewritten in terms of the orthogonal pair $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, so that it can be factorised into two multiplicative terms $L_1(\boldsymbol{\alpha})$ and $L_2(\boldsymbol{\beta})$. Each term depends either on $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$, but not on both. Clearly, if our interest is in $\boldsymbol{\alpha}$ but not in $\boldsymbol{\beta}$, it is sufficient to only consider $L_1(\boldsymbol{\alpha})$ and to ignore $L_2(\boldsymbol{\beta})$. The term $L_1(\boldsymbol{\alpha})$ is precisely the product of the conditional probability mass functions of $y_{ij1}|y_{ij\bullet}$. Thus estimating $\boldsymbol{\alpha}$ by fitting the logistic regression model (4.1) is valid.

We now turn to the case of more than two strata, i.e. $R > 2$. Let $\boldsymbol{y}_{ij}$ denote the vector $(y_{ij1}, \ldots, y_{ijR})^T$ and $y_{ij\bullet} = y_{ij1} + \ldots + y_{ijR}$ denote the corresponding sum. The conditional distribution of $\boldsymbol{y}_{ij}$, given $y_{ij\bullet}$, is then multinomial with individual success probabilities

$$\begin{aligned}
\pi_{ijr} &= \frac{n_{ijr}\lambda_{ijr}}{\sum_{s=1}^{R} n_{ijs}\lambda_{ijs}} \\
&= \frac{n_{ijr}/n_{ijR} \cdot \lambda_{ijr}/\lambda_{ijR}}{1 + \sum_{s=1}^{R-1}(n_{ijs}/n_{ijR} \cdot \lambda_{ijs}/\lambda_{ijR})} \\
&= \frac{\exp\left(\log\left(\frac{n_{ijr}}{n_{ijR}}\right) + \Delta_\mu^{(r)} + \Delta_j^{(r)} + \Delta_k^{(r)}\right)}{1 + \sum_{s=1}^{R-1} \exp\left(\log\left(\frac{n_{ijs}}{n_{ijR}}\right) + \Delta_\mu^{(s)} + \Delta_j^{(s)} + \Delta_k^{(s)}\right)}
\end{aligned}$$

for $r = 1, \ldots, R-1$. The success probability π_{ijR} is implicitly given by $1 - \sum_{s=1}^{R-1} \pi_{ijs}$. Here $\Delta_\mu^{(r)} = \mu_r - \mu_R$, $\Delta_j^{(r)} = \varphi_{jr} - \varphi_{jR}$, and so on. This can be identified as a multinomial logistic regression model with offsets $\log(n_{ijr}/n_{ijR})$, $r = 1, \ldots, R-1$, see e.g. McCullagh and Nelder (1989, Chapter 5) or Fahrmeir and Tutz (2001, Section 3.2). A conditional likelihood argument can be used in the same way as above to show that this multinomial likelihood carries the same information about the relative risk parameters as the original Poisson likelihood. Thus the unconditional analysis with a Poisson regression model and the conditional analysis using

bi- or multinomial logistic regression will lead to the same ML estimates of (average) relative risk.

However, for the case in which overdispersion is accounted for differences are possible. This can be explained by recognising that overdispersion is often caused by additional unobserved variables z_{ijr}, say, which act additively on the linear predictor (Breslow, 1984). The variance of the z_{ijr}'s represents the amount of overdispersion. Suppose now that, for each i and j, such unobserved explanatory variables act in fact *simultaneously* on the different strata, i.e. $z_{ij1} = z_{ij2} = z_{ij3}$. In the calculation of the relative risk parameters based on the difference of the extended linear predictor, the z_{ijr}'s would then cancel and overdispersion would be removed. Of course, identical unobserved explanatory variables across strata are implausible. However, in applications it may often be the case that the z_{ijr}'s are positively correlated across strata, if the structural part of the model is specified adequately. The variance of the difference of the z_{ijr}'s would then be smaller and overdispersion will be reduced in the conditional approach. Extra-Poisson variation induced by correlated unobserved explanatory variables can thus be reduced through conditioning. Indeed, in the application of Section 6 we observed that applying the conditional approach to well-fitting models, i.e. models in which the structural part is specified appropriately, overdispersion is reduced.

A further advantage of the conditional approach is that it ignores nuisance parameters which are not of interest anyway. It estimates directly the log relative risk parameters, suitably smoothed if appropriate. This is in contrast to the unconditional formulation, where the original age, period and cohort parameters in the different strata are often smoothed (Knorr-Held and Rainer, 2001; Bray, 2002). The number of time effects in the conditional formulation is hence smaller than in the original formulation. This is a considerable simplification, since the number of smoothing terms is reduced. For example, in the Apc model with $R = 2$ strata, only the difference in stratum-specific cohort effects and perhaps also the difference in stratum-specific period effects need to be smoothed, rather than the original period and cohort effects in the two strata and the common age effect.

In the following applications we estimate the multinomial logistic regression models using the R-package `VGAM` (Yee, 2009, version 0.7-9). The approach allows for quasi-likelihood adjustments for overdispersion based on Pearson's statistic X^2 for multinomial data (McCullagh and Nelder, 1989, Section 5.5). Here the standard errors of the parameter estimates are inflated by the square root of an estimated overdispersion factor $\hat{\phi}_{\mathrm{C}}$, equal to X^2 divided by the residual degrees of freedom. In the absence of overdispersion, Akaike Information Criterion (AIC) can be computed for model choice. Adjustments of AIC for a quasi-likelihood approach exist, for example the so-called QAIC criterion (Burnham and Anderson, 2002). QAIC is defined as

$$\mathrm{QAIC} = -2 \log L / \hat{\phi}_{\mathrm{C}}^{apc} + 2p$$

where L is the value of the maximised likelihood function, p is the number of parameters and $\hat{\phi}_{\mathrm{C}}^{apc}$ is the estimated overdispersion parameter from the apc model, the largest model available (Pan, 2001; Burnham and Anderson, 2002). If an overdispersion factor is estimated, then one additional parameter must be added to p.

An attractive feature of the `VGAM` package is that it allows to smooth the parameter estimates using cubic smoothing splines. For the COPD data introduced in Section 2, smoothing the cohort effects is of primary interest, since there are 110 cohort parameters and the unsmoothed ML estimates tend to be quite irregular. Of course, age or period effects could also be smoothed, if necessary. Note that the Apc model induces an additional identifiability problem if the width of the age and period intervals is not equal (Holford, 2006; Riebler and Held, 2010). Smoothing of the period or cohort effects is then essential to avoid artificial periodicities in the parameter estimates. Example code for the conditional estimation of APC models for multiple outcomes with the package `VGAM` is given on the website `http://www.biostat.uzh.ch/research/Rpacks_en.html`. Adjustments for overdispersion and the fit of cubic smoothing splines are illustrated.

5 Extension to data with more stratification levels

Suppose now that the data are further stratified by a variable $g = 1, \ldots, G$, for example gender. Counts y_{ijrg} are then available with corresponding population counts n_{ijrg}. We now propose an extension of the ordinary MAPC model to account for the additional stratification variable. The general construction is as follows. Those effects which do already depend on the original stratification variable r are kept in the model as before. However, the effects assumed to be identical across the original strata are allowed to depend on the new stratification variable g. In addition, we let the intercept depend on both r and g.

For example, consider the Apc model (3.2). An extended formulation is

$$\eta_{ijrg} = \mu_{rg} + \theta_{ig} + \varphi_{jr} + \psi_{kr}, \tag{5.1}$$

say, i.e. the age effects θ_{ig} are allowed to vary across the new stratification variable whereas period and cohort effects remain to vary across the original stratification variable. Of course, other combinations are possible, for example period effects which do depend on g but not on r. Note that the formulation is indeed more general than the original MAPC model, where at least one time effect had to be identical across all strata. Now every time effect can vary according to one of the two strata.

Suitable conditioning allows to remove either the time effects depending on r, or the remaining time effects depending on g. Thus, two separate multinomial logistic regression models can be fitted. For example, in model (5.1) conditioning on $y_{ijr\bullet}$ will enable to estimate the difference in stratum-specific age effects, whereas conditioning on $y_{ij\bullet g}$ will be useful to estimate differences in stratum-specific period and cohort effects. Note that the difference in stratum-specific age effects does compare age effects in different strata of the new variable, whereas the differences in stratum-specific period and cohort effects relate to the original strata $r = 1, \ldots, R$. The conditional approach in this specific model will be discussed in more detail in Section 6.2.

Incidentally, a third stratification variable $h = 1, \ldots, H$ could also be incorporated with age

effects depending on g, period effects depending on h and cohort effects depending on r, say:

$$\eta_{ijrgh} = \mu_{rgh} + \theta_{ig} + \varphi_{jh} + \psi_{kr}. \tag{5.2}$$

Conditioning on $y_{ij\bullet gh}$, $y_{ijr\bullet h}$, and $y_{ijrg\bullet}$, respectively, would then allow to estimate the differences $\Delta_k^{(r)} = \psi_{kr} - \psi_{kR}$, $\Delta_i^{(g)} = \theta_{ig} - \theta_{iG}$, and $\Delta_j^{(h)} = \varphi_{jh} - \varphi_{jH}$ using three separate multinomial logistic regression models.

6 Application: COPD mortality in England & Wales

We begin with a detailed analysis of males only. Then we consider an extended MAPC model to analyse males and females jointly.

6.1 Analysis of heterogeneous time trends across regions

Different MAPC models have been applied to the data on males using the conditional formulation proposed in this paper. Table 1 gives the estimated overdispersion parameter $\hat{\phi}_{\mathrm{C}}$ and the QAIC value for model comparison. The apc model, see (3.1), is listed for comparison since it is the largest model available. The estimated overdispersion parameter from this model is used in the calculation of QAIC for all models. Of course, the estimates from the apc model suffer from the usual identifiability problem in univariate APC models and cannot be displayed. Among the three models with only one relative risk effect varying across strata, the ApC model with stratum-specific period effects is considered the best model. Not taking into account the apc model for which relative risks are not identifiable the model allowing period and cohort effects (Apc) to differ across strata is clearly the best among all models. The findings are similar to those obtained from an unconditional analysis using a Bayesian predictive model selection criterion (Riebler and Held, 2010, Table 3).

For comparison, Table 1 lists also the estimated overdispersion $\hat{\phi}_{\mathrm{UC}}$ from an unconditional analysis by ML using age, period and cohort effects as simple factor variables. For the models

Table 1: QAIC from the conditional approach and dispersion parameters ϕ_{C} and ϕ_{UC} of both the conditional and unconditional approach for all models. Shown is the difference of QAIC and $\mathrm{QAIC}_{\mathrm{apc}} = 4247.072$. The apc model is the largest model; its estimated dispersion parameter ϕ_{C}^{apc} is used to calculate the QAIC values for all models. The last row gives the reduction of the adjusted standard errors in the conditional approach, relative to the unconditional approach.

	APC	aPC	ApC	APc	apC	aPc	Apc	apc
QAIC	4638.15	2643.08	2085.02	2743.95	268.09	906.56	80.07	0.00
$\hat{\phi}_{\mathrm{C}}$	9.28	6.02	5.32	6.14	1.81	2.98	1.34	1.18
$\hat{\phi}_{\mathrm{UC}}$	33.30	5.48	4.89	6.17	2.24	3.01	2.10	1.91
$\sqrt{\hat{\phi}_{\mathrm{C}}/\hat{\phi}_{\mathrm{UC}}}$	0.53	1.05	1.04	1.00	0.90	0.99	0.80	0.79

with a relatively poor fit in terms of QAIC, the estimates of $\hat{\phi}_{\mathrm{C}}$ and $\hat{\phi}_{\mathrm{UC}}$ are quite similar. The only exception is the (clearly inappropriate) APC model, where $\hat{\phi}_{\mathrm{C}}$ is much smaller than $\hat{\phi}_{\mathrm{UC}}$. For the two best fitting models (Apc and apc), however, $\hat{\phi}_{\mathrm{C}}$ is also considerably smaller than $\hat{\phi}_{\mathrm{UC}}$. The overdispersion parameter is used in both approaches to inflate the estimated variance of the relative risk estimates. Without such adjustments, estimates and standard errors from the unconditional and the conditional analysis must be identical, due to the arguments described in Section 4. However, the adjusted standard errors in the Apc model from the conditional analysis are 20% smaller than in the unconditional analysis ($\sqrt{\hat{\phi}_{\mathrm{C}}/\hat{\phi}_{\mathrm{UC}}} = 0.80$). The conditional analysis hence leads to relative risk estimates of higher precision. As suggested in Section 4 this might be, because the conditional approach can handle to some extent unmeasured confounding.

We have been able to confirm this hypothesis through an exploratory analysis of the Pearson residuals res_{ijr}, obtained from the unconditional ML analyses. Table 2 gives the pairwise correlations between all stratum-specific Pearson residuals $\mathbf{res}_r$, $r = 1, 2, 3$. With the exception of the APC model, positive correlations for all three pairs of strata can be observed exactly for those models where overdispersion has been reduced by 20% or more in the conditional analysis (Apc and apc model).

Turning to the actual parameter estimates, Figure 1 shows the results of all MAPC models when allowing one of age, period or cohort effects to vary across strata. The estimated relative

Table 2: Correlation of stratum-specific Pearson residuals in the unconditional ML approach.

	APC	aPC	ApC	APc	apC	aPc	Apc	apc
Corr($\mathbf{res}_1$,$\mathbf{res}_2$)	0.291	0.305	-0.552	-0.537	0.233	0.009	0.198	0.258
Corr($\mathbf{res}_2$,$\mathbf{res}_3$)	-0.783	-0.416	0.141	-0.282	0.174	0.178	0.341	0.453
Corr($\mathbf{res}_1$,$\mathbf{res}_3$)	-0.610	-0.583	-0.177	0.132	-0.129	-0.185	0.217	0.254

risk obtained from the APc model is based on a model with cubic splines for the differences in stratum-specific cohort effects. We used cubic smoothing splines with knots placed in ten-year intervals. This corresponds to the periodicity induced by the identifiability problem described in detail in Holford (2006). Knots at higher indices than 70 were found to induce a non-interpretable pattern at cohorts with a centre later than 1960 due to the limited amount of data available for younger cohorts. Alternatively, regression splines constrained to be linear in the tails have been proposed in Heuer (1997). For comparison, the estimates based on a model without splines are also given. It is clear that smoothing of the difference in cohort effects is important to remove spurious random variation in the ML estimates, especially for younger cohorts. In the ApC model we have decided not to smooth the difference in period effects since there was enough data available to obtain reliable estimates without smoothing. In this application there is particular interest in year-to-year variation in COPD mortality (Hansell *et al.*, 2003) and smoothing may blur these time trends to some extent.

Figure 2 illustrates that unsmoothed relative risk estimates of period and cohort effects in the Apc model, even when using the conditional approach, show artificial cyclical patterns which make them not interpretable, see Riebler and Held (2010). The problem is resolved by smoothing the cohort relative risks (Figure 3). As before, we used cubic smoothing splines with knots placed in ten-year intervals. The estimates are very similar to those obtained using a fully Bayesian approach based on the original Apc model, see Figure 4.

We now briefly interpret the relative risk estimates, for more details see Riebler and Held (2010). The observed higher average relative risk $\exp(\tilde{\Delta}_j^{(2)})$ of conurbations compared to $\exp(\tilde{\Delta}_j^{(1)})$ of

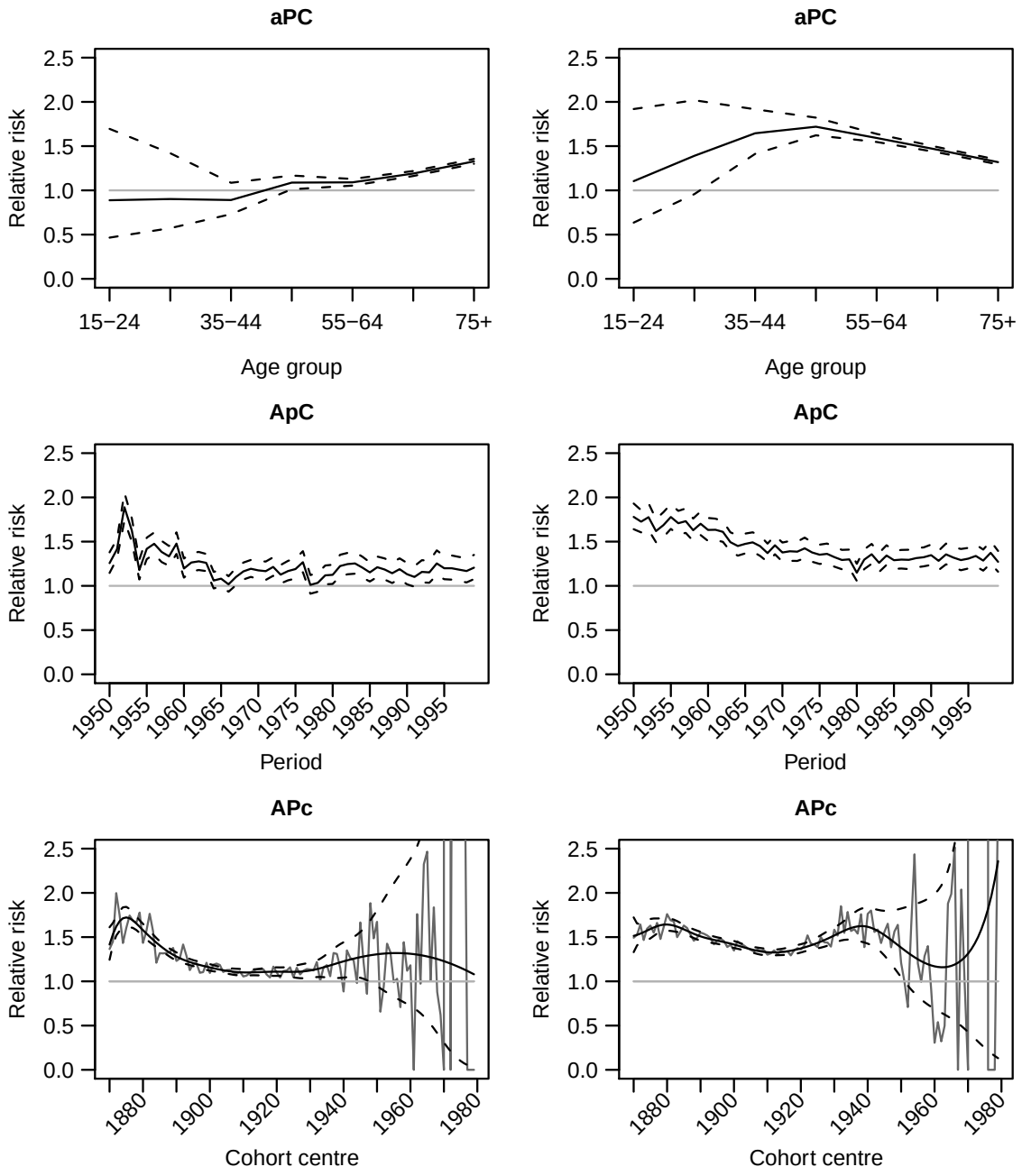

Figure 1: Results of the conditional approach analysing the aPC, ApC and APc models. The left column shows the relative risks (ML estimates within 95% confidence intervals) of Greater London to rural areas, the right column those of conurbations excluding Greater London to rural areas. Each row corresponds to one model. In the last row (APc model), the grey lines represent the estimates based on a spline-free model.

Greater London can be explained by the fact that conurbations are mainly situated in northern England, where the former predominance of heavy industries resulted in especially high levels of air pollution. The colder climate in the North might also be related (Law and Morris, 1998; Hansell, 2004). There is substantial year-to-year variation in the estimated average relative risk of period effects from 1950 ($j = 1$) to 1999 ($j = 50$) with higher values in years of known air pollution episodes. For example, the increased average relative risk in 1952 is probably related to the 1952 "Great Smog" in London. Some peaks may also be caused by influenza epidemics (Hansell, 2004; Wedzicha, 2004). The average relative risk of cohort effects differs remarkably between Greater London and conurbations. Possible reasons for the shifted occurrence of pronounced estimates could be different smoking behaviour in these two strata.

6.2 Analysis of heterogeneous time-trends across regions and gender

Figure 5 gives estimates from a separate application of the Apc model to the corresponding data on females. The estimates are very similar to the corresponding ones on males (Figure 3) so a joint analysis of COPD mortality for males and females may be of interest. Let g be an indicator for gender with $g = ♂$ for males and $g = ♀$ for females and let $r = 1, 2, 3$ denote the strata Greater London, conurbations excluding Greater London and rural areas. A potentially useful model would allow age effects to be different across gender, while period and cohort effects will be different across regions. Assuming a gender and region dependent intercept μ_{rg} the corresponding linear predictor is as in (5.1). Females and rural areas are chosen in the following as reference categories.

To estimate the difference in gender-specific age effects $\Delta_i = \theta_{i♂} - \theta_{i♀}$ of males to females and the difference in region-specific period and cohort effects $\Delta_j^{(r)} = \varphi_{jr} - \varphi_{j3}$ and $\Delta_k^{(r)} = \psi_{kr} - \psi_{k3}$ of Greater London and conurbations ($r = 1, 2$) relative to rural areas ($r = 3$), two separate conditional analyses are necessary. In the first we condition on the sum $y_{ijr\bullet} = y_{ijr♂} + y_{ijr♀}$ across gender to eliminate period and cohort effects. In the second we condition on the sum $y_{ij\bullet g} = y_{ij1g} + y_{ij2g} + y_{ij3g}$ across regions to eliminate the age effects.

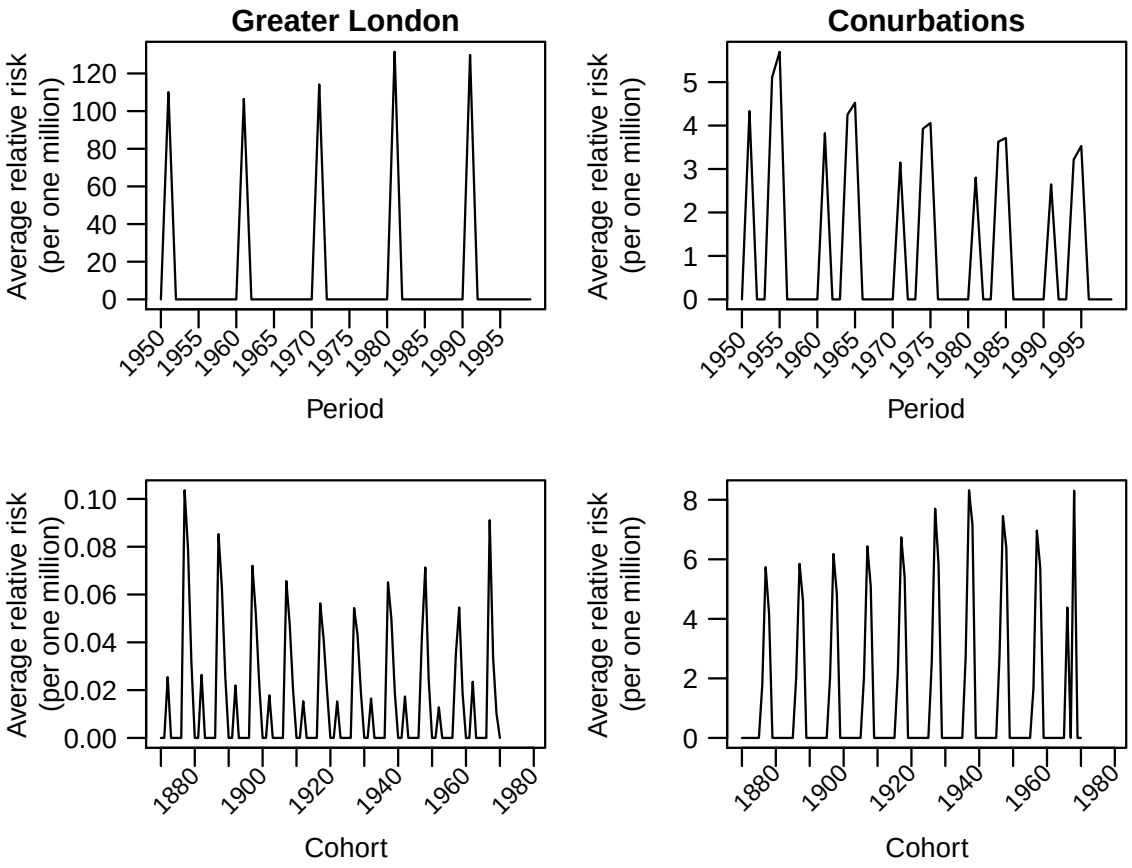

Figure 2: Results of the spline-free Apc model. The left column shows the average relative risk estimates of Greater London to rural areas, the right column those of conurbations excluding Greater London to rural areas.

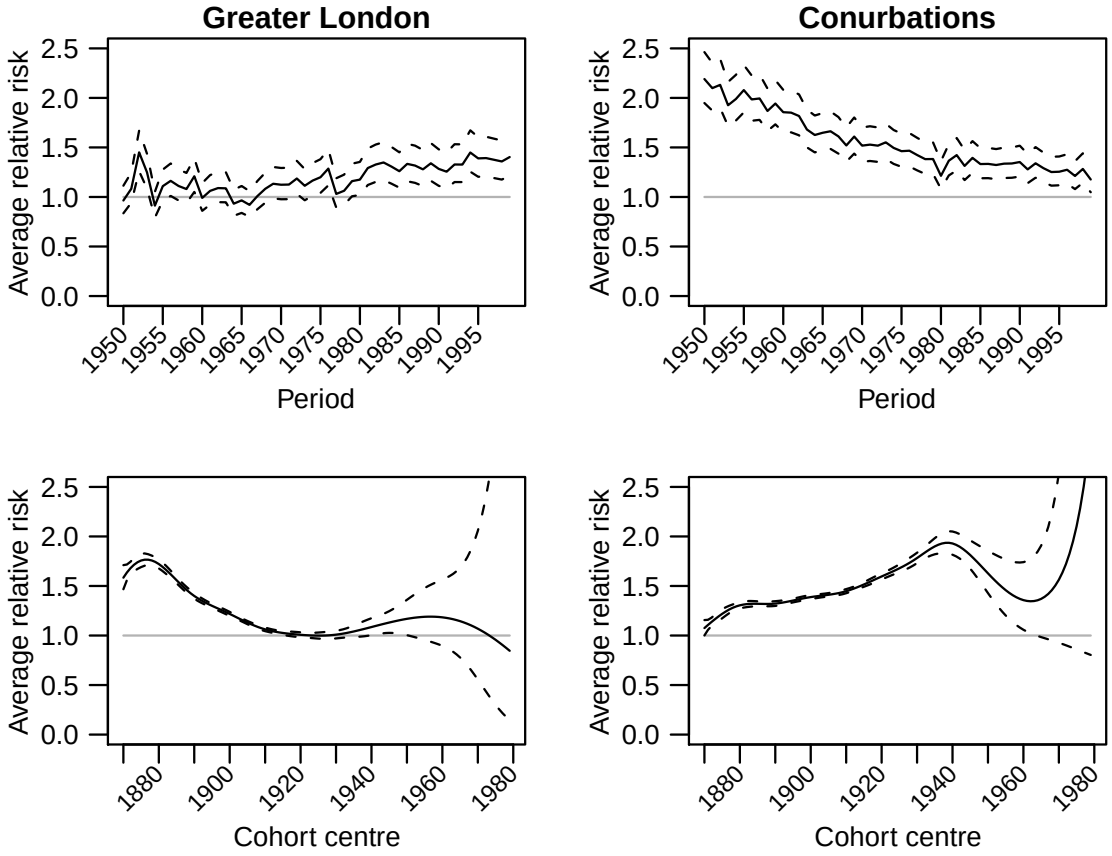

Figure 3: Results of the Apc model for males obtained by smoothing the difference of cohort effects using a cubic spline. The left column shows the average relative risks (ML estimates within 95% confidence intervals) of Greater London to rural areas, the right column those of conurbations excluding Greater London to rural areas.

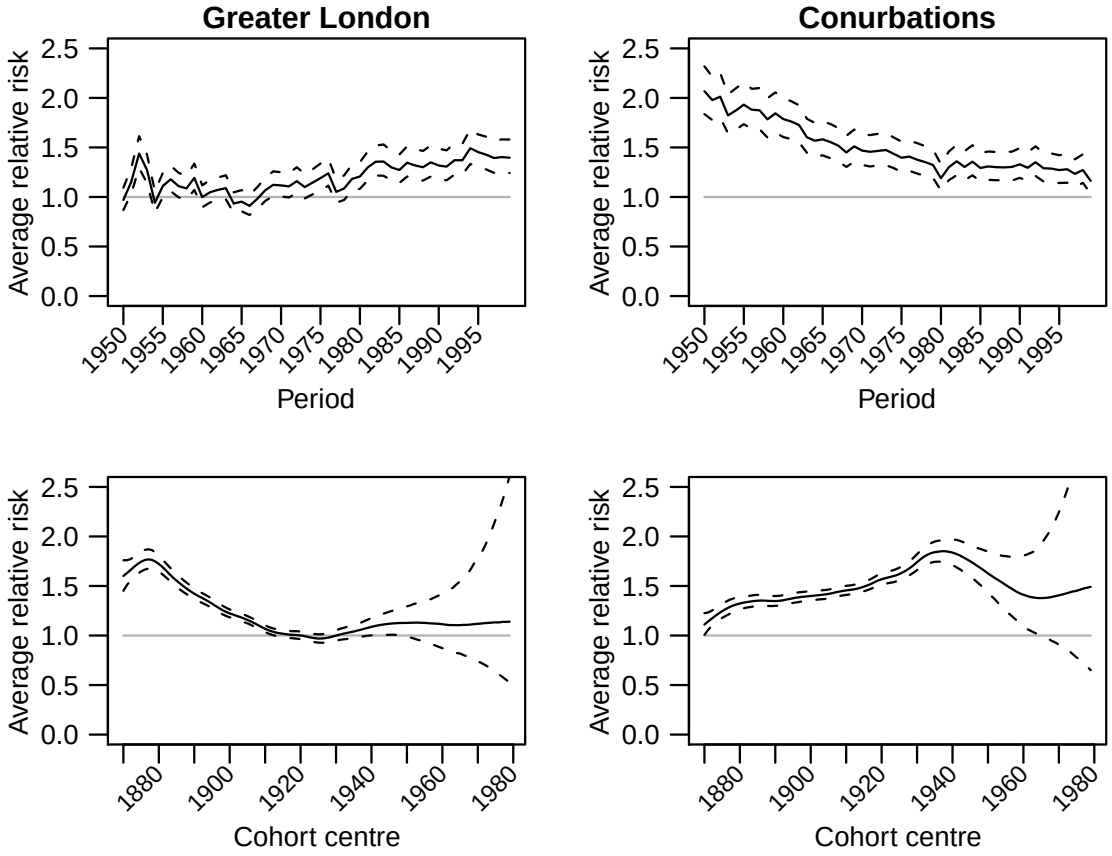

Figure 4: Results of the Apc model for males obtained by an unconditional Bayesian approach. The left column shows the average relative risks (median estimates within 95% credibility intervals) of Greater London to rural areas, the right column those of conurbations excluding Greater London to rural areas.

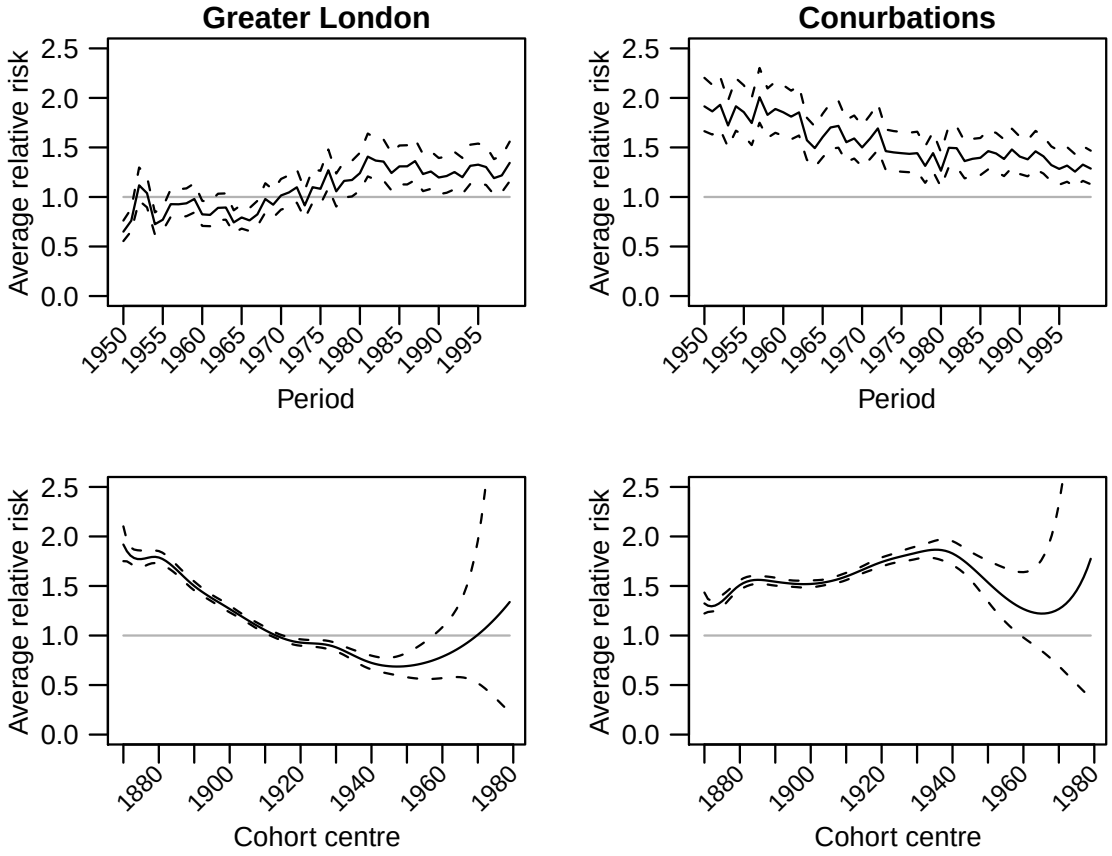

Figure 5: Results of the Apc model for females obtained by smoothing the difference of cohort effects using a cubic spline. The left column shows the average relative risks (ML estimates within 95% confidence intervals) of Greater London to rural areas, the right column those of conurbations excluding Greater London to rural areas.

For the first approach, it is easy to see that $y_{ijr♂}|y_{ijr\bullet}$ is binomial with success probability

$$\pi_{ijr♂} = \text{expit}\left(\log\frac{n_{ijr♂}}{n_{ijr♀}} + \Delta_{\mu,r} + \Delta_i\right) \tag{6.1}$$

for $r \in \{1,2,3\}$. Here $\Delta_{\mu,r} = \mu_{r,♂} - \mu_{r,♀}$. Thus the difference in gender-specific age effects Δ_i can be estimated through a logistic regression model of the form (6.1), applied to all three regions $r = 1,2,3$ with region-specific intercepts $\Delta_{\mu,r}$.

Similarly, if we condition on $y_{ij\bullet g}$, the vector $\boldsymbol{y}_{ijg} = (y_{ij1g}, y_{ij2g}, y_{ij3g})^T$ is trinomial distributed with success probabilities

$$\pi_{ijrg} = \frac{\exp\left(\log\left(\frac{n_{ijrg}}{n_{ij3g}}\right) + \Delta^{(r)}_{\mu,g} + \Delta^{(r)}_j + \Delta^{(r)}_k\right)}{1 + \sum_{s=1}^{2}\exp\left(\log\left(\frac{n_{ijsg}}{n_{ij3g}}\right) + \Delta^{(s)}_{\mu,g} + \Delta^{(s)}_j + \Delta^{(s)}_k\right)} \tag{6.2}$$

for $r = 1,2$ and $g \in \{♂,♀\}$. Here $\Delta^{(r)}_{\mu,g} = \mu_{r,g} - \mu_{3,g}$. As before, the success probability π_{ij3g} is given by $1 - \pi_{ij1g} - \pi_{ij2g}$. Therefore the difference in region-specific period and cohort effects $\Delta^{(r)}_j$ and $\Delta^{(r)}_k$ can be estimated by a trinomial logistic regression model of the form (6.2), applied to both males and females with gender- and region-specific intercepts $\Delta^{(r)}_{\mu,g}$.

Figure 6 shows in the top row the corresponding relative risk $\exp(\Delta_{\mu,r} + \Delta_i)$ of males to females for Greater London ($r = 1$) and conurbations excluding Greater London ($r = 2$). With increasing age the risk for a male to die of COPD increases to nearly four times that of a female in Greater London, with a slightly lower risk for conurbations. In the two youngest age groups, the risk is approximately equal. This pattern can be explained by the higher smoking prevalence of males. Note that the estimates for each region differ only by the multiplicative factor $\exp(\Delta_{\mu,r})$. The corresponding estimates for rural areas ($r = 3$, not shown) are very similar to those obtained for Greater London.

The second row in Figure 6 shows the average relative risk $\exp(\Delta^{(r)}_{\mu,g} + \Delta^{(r)}_j)$ in year j for strata $r = 1,2$ relative to rural areas ($r = 3$) and for both males and females. The last row gives the corresponding average relative risks in cohort k. The relative risks for Greater London

compared with rural areas show virtually no difference across gender. However the relative risk for conurbations compared with rural areas is somewhat higher for females. The estimated patterns are very similar to those obtained from a separate analysis of males and females (Figure 3 and 5).

7 Concluding Remarks

The conditional formulation presented in this paper simplifies the analysis and interpretation of multivariate age-period-cohort models. One advantage of the new approach is that it models directly the parameters of interest, namely the (average) relative risks, whereas the unconditional approach includes a number of nuisance parameters not identifiable through the likelihood. Additionally, the conditional approach can handle to some extent unmeasured confounding. Reduced overdispersion and relative risk estimates with smaller standard errors are the consequence, as exemplified in our application. For model estimation standard software for multinomial logistic regression can be used.

We have also proposed an extended MAPC model framework to analyse data stratified by more than one variable. The conditional approach is still applicable and leads to separate multinomial logistic regression models which enable the estimation of relative risks in the different strata. This has been illustrated through an analysis of COPD mortality in England & Wales stratified by geographical region and gender. In principle an unconditional approach using MCMC as presented in Riebler and Held (2010) could be developed for data with more than one stratification variable as well. However, this would require considerable additional programming effort whereas the conditional approach is much simpler to apply.

If appropriate, the difference in stratum-specific time trends obtained from the conditional formulation (rather than the original age, period and cohort parameters in the original formulation) can be smoothed. However, complications of the Apc model due to unequal intervals of age groups and periods are inherited from the unconditional formulation. Smoothing the difference in region-specific cohort effects with cubic smoothing splines resolved the identifiability

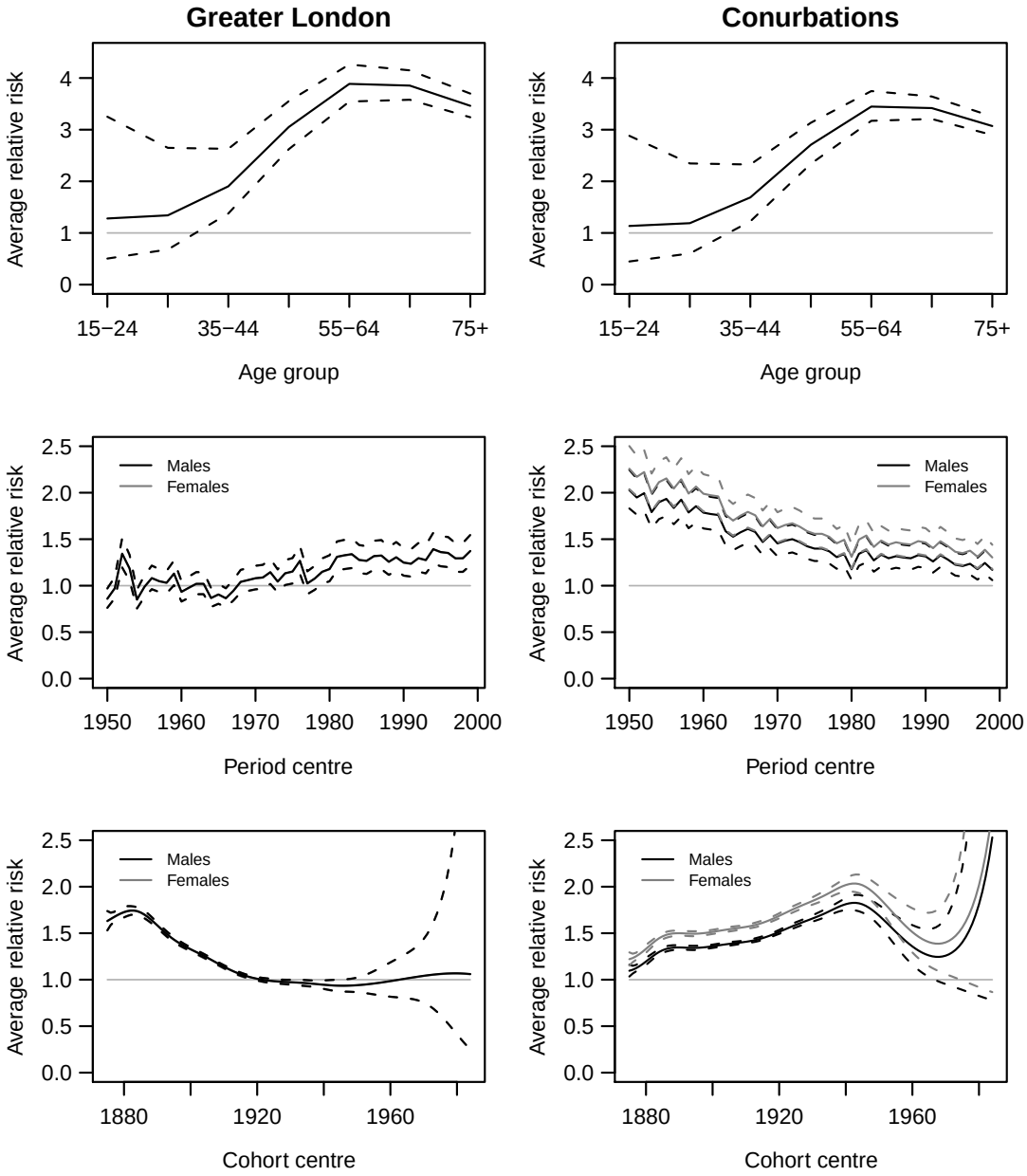

Figure 6: Average relative risk (ML estimates within 95% confidence intervals) of COPD death. The top row gives risk estimates for males relative to females in Greater London (left) and conurbations excluding Greater London (right). The remaining figures give average period (second row) and cohort (bottom row) risk estimates of Greater London (left) and conurbations excluding Greater London (right), relative to rural areas, separately for males and females.

problem in this case.

Of course, a smoothing approach which automatically estimates the smoothing parameters (here implicitly specified by the number and location of the knots) may be preferable. For example, penalised splines (P-splines) would offer this possibility, either in a frequentist (Eilers and Marx, 1996) or Bayesian framework (Lang and Brezger, 2004; Brezger and Lang, 2006). Alternatively, second-order random walk priors with unknown variances could be used (Besag *et al.*, 1995). A fully Bayesian approach typically requires MCMC for inference, although approximate inference based on the integrated nested Laplace approximation (INLA) (Rue *et al.*, 2009) avoids this. However, only the binomial, not the multinomial model is currently implemented in the `inla` program (`www.r-inla.org`). The P-spline approach combined with a multinomial logistic regression model is available in `BayesX` (Brezger *et al.*, 2005) within either a fully or empirical Bayes framework. However, adjustments for overdispersion, either based on correlated normal random effects (Aitchison and Shen, 1980) or using the Dirichlet-multinomial distribution (Poortema, 1999), are difficult to implement and currently not available in `BayesX`. We plan to consider such extensions in future work. For example, through applying the multinomial-Poisson transformation (Baker, 1994) inference based on a Poisson likelihood might be possible within the `inla` program. Model-based adjustments for overdispersion by including additional random effects are then straightforward.

Acknowledgements

This work was supported by the Swiss National Science Foundation. The authors express thanks to Anna Hansell for access to the dataset on COPD mortality in England & Wales. Several helpful comments by the editor and two anonymous referees are gratefully acknowledged.

References

Agresti, A. (2002). *Categorical Data Analysis*, 2nd edn, John Wiley & Sons, Ltd, Chichester.

Aitchison, J. and Shen, S. M. (1980). Logistic-normal distributions: some properties and uses, *Biometrika* **67**: 261–272.

Baker, S. G. (1994). The multinomial-Poisson transformation, *Statistician* **43**(4): 495–504.

Berzuini, C. and Clayton, D. (1994). Bayesian analysis of survival on multiple time scales, *Statistics in Medicine* **13**(8): 823–838.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion), *Statistical Science* **10**: 3–66.

Bray, I. (2002). Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **51**(2): 151–164.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models, *Applied Statistics* **33**: 38–44.

Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on Bayesian P–splines, *Computational Statistics and Data Analysis* **50**: 967–991.

Brezger, A., Kneib, T. and Lang, S. (2005). BayesX: Analyzing Bayesian structural additive regression models, *Journal of Statistical Software* **14**(11): 1–22.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn, Springer, New York.

Casella, G. and Berger, R. L. (1990). *Statistical Inference*, Duxbury Press, Belmont, California.

Clayton, D. and Hills, M. (1993). *Statistical Models in Epidemiology*, Oxford University Press, Oxford.

Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models, *Statistics in Medicine* **6**(4): 469–481.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties, *Statistical Science* **11**(2): 89–102.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd edn, Springer–Verlag, New York.

Fu, W. J. (2000). Ridge estimator in singular design with application to age-period- cohort analysis of disease rates, *Communications in Statistics - Theory and Method* **29**: 263–278.

Gelfand, A. E. and Sahu, K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models, *Journal of the Americal Statistical Association* **94**(445): 247–253.

Hansell, A., Knorr-Held, L., Best, N., Schmid, V. and Aylin, P. (2003). COPD mortality trends 1950-1999 in England & Wales - Did the 1956 Clean Air Act make a detectable difference?, *Epidemiology* **14**(5): S55.

Hansell, A. L. (2004). *The Epidemiology of Chronic Obstructive Pulmonary Disease in the UK: Spatial and Temporal Variations*, PhD thesis, Faculty of Medicine of the University of London, Imperial College, St Mary's Campus.

Heuer, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines, *Biometrics* **53**(1): 161–177.

Hilbe, J. M. (2007). *Negative Binomial Regression*, Cambridge University Press, Cambridge.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates, *Biometrics* **39**: 311–324.

Holford, T. R. (1991). Understanding the effects of age, period, and cohort on incidence and mortality rates, *Annual Review of Public Health* **12**: 425–457.

Holford, T. R. (1998). Age-period-cohort analysis, *in* P. Armitage and T. Colton (eds), *Encyclopedia of Biostatistics*, Wiley: Chichester, pp. 82–99.

Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals, *Statistics in Medicine* **25**: 977–993.

Jacobsen, R., von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different?, *European Journal of Epidemiology* **19**: 117–121.

Kelsall, J. and Diggle, P. (1998). Spatial variation in risk: a nonparametric binary regression approach, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47**: 559–573.

Knorr-Held, L. and Rainer, E. (2001). Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction, *Biostatistics* **2**(1): 109–129.

Lang, S. and Brezger, A. (2004). Bayesian P–splines, *Journal of Computational and Graphical Statistics* **13**: 183–212.

Law, M. and Morris, J. (1998). Why is mortality higher in poorer areas and in more northern areas of England and Wales?, *Journal of Epidemiology and Community Health* **52**(6): 344–352.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Chapman and Hall, London.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations, *Biometrics* **57**: 120–125.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press, Oxford.

Poortema, K. (1999). On modelling overdispersion of counts, *Statistica Neerlandica* **53**: 5–20.

Riebler, A. and Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models, *Biostatistics* **11**: 57–69.

Roberts, G. O., Sahu, S. K. and Gilks, W. R. (1995). Comment on "Bayesian Computation and Stochastic Systems", by J. Besag, P. Green, D. Higdon, and K. Mengersen, *Statistical Science* pp. 49–51.

Robertson, C. and Boyle, P. (1998). Age-period-cohort analysis of chronic disease rates. I: Modelling approach., *Statistics in Medicine* **17**(12): 1305–1323.

Robertson, C., Gandini, S. and Boyle, P. (1999). Age-period-cohort models: A comparative study of available methodologies, *Journal of Clinical Epidemiology* **52**(6): 569–583.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**: 319–392.

Wedzicha, J. A. (2004). Role of viruses in exacerbations of chronic obstructive pulmonary disease, *Proceedings of the American Thoracic Society* **1**(2): 115–120.

Yang, Y., Schulhofer-Wohl, S., Fu, W. J. and Land, K. C. (2008). The intrinsic estimator for age-period-cohort analysis: What it is and how to use it, *American Journal of Sociology* **113**(6): 1697–1736.

Yee, T. W. (2009). *A User's Guide to the* `VGAM` *Package*. Available from: `http://www.stat.auckland.ac.nz/~yee/VGAM`.

Zheng, T. Z., Holford, T. R., Chen, Y. T., Ma, J. Z., Flannery, J., Liu, W., Russi, M. and Boyle, P. (1996). Time trend and age-period-cohort effect on incidence of thyroid cancer in Connecticut, 1935-1992, *International Journal of Cancer* **67**(4): 504–509.

PAPER III

Correlated multivariate age-period-cohort models

Andrea Riebler, Leonhard Held & Håvard Rue

Technical Report, University of Zurich.

Correlated multivariate age-period-cohort models

Andrea Riebler[1*], Leonhard Held[1], Håvard Rue[2]

[1] Biostatistics Unit, Institute of Social and Preventive Medicine,

University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland;

Email: {andrea.riebler,leonhard.held}@ifspm.uzh.ch,

[2] Department of Mathematical Sciences, Norwegian University of Science and Technology,

N-7491 Trondheim, Norway; Email: havard.rue@math.ntnu.no

Multivariate age-period-cohort models have recently been proposed for the analysis of heterogeneous time trends. For a fully Bayesian analysis Gaussian Markov random field (GMRF) priors are typically used. However, standard GMRF priors do not account for a potential dependence between outcomes. We present an extended approach based on correlated smoothing priors and correlated overdispersion parameters. This approach is useful to improve the precision of relative risk estimates and to extrapolate and forecast missing data. Algorithmic routines are based on either Markov chain Monte Carlo or integrated nested Laplace approximations. The methodology is applied to chronic obstructive pulmonary disease mortality of males in England & Wales and to overall mortality of females in Scandinavia. For model comparison we use the deviance information criterion (DIC), proper scoring rules and the log marginal likelihood.

Keywords: Bayesian analysis; Gaussian Markov random field; INLA; Multivariate age-period-cohort model; Uniform correlation matrix.

*To whom correspondence should be addressed.

1 Introduction

Projections of mortality or incidence data are important for the planning of public health measures. For historical reasons projections into the past are interesting. Frequently, age-period-cohort (APC) models are used to analyse and project data stratified by age group and period. APC models analyse age-specific vital rates according to three time scales: age at diagnosis (age), date of diagnosis (period) and date of birth (cohort). The linear dependence of age, period and cohort induces the well-known identifiability problem, however predictions based on APC models are uniquely determined and not affected (Holford, 1985).

To make projections, future values of the time effects, i.e. age, period and cohort effects, are needed. Using classical maximum-likelihood (ML) approaches, predictions often rely on strong parametric assumptions. For example, Osmond (1985) assumed constant age effects, and obtained future period and cohort effects by a linear regression applied to a chosen number of most recent effects on each scale. This method strongly depends on the chosen number of past values and the type of regression applied (i.e. weighted or unweighted) (Bray, 2002). Bayesian versions of the APC model facilitate prediction through the assumed smoothness of age, period and cohort effects which can be extrapolated into both the future and past (Knorr-Held and Rainer, 2001). A natural choice is to use second-order random walks (RW2s) as smoothing prior, as they penalise the second differences which are identifiable from the likelihood.

So far, mortality or incidence data were mostly projected at the univariate level. However, frequently age-specific rates may exist with a comparable progression, for example data from a neighbouring country, the opposite sex or a related disease. Simultaneous modelling of these rates can be valuable if the considered rates share risk factors. The precision of the projections obtained via an univariate APC model may be considerably improved if the outcome-specific rates are correlated.

For the case in which age-specific rates are available for different strata, multivariate APC models have been proposed, see e.g. Jacobsen *et al.* (2004) or Riebler and Held (2010). Multivariate APC models share sets of parameters, frequently the age effects, while the effects of the

remaining time scales, here period and cohort effects, may differ across strata. Bayesian formulations are particularly attractive for estimating relative risks, since they avoid severe artefacts of classical ML techniques through the use of smoothing priors. Furthermore, they can easily account for unobserved heterogeneity (overdispersion) by introducing additional random effect parameters. Typically, both overdispersion parameters and smoothing priors on the time trends are assumed to be independent across strata. Hence, a potential dependence between them, caused, for example, by the same risk factors, is not captured.

We propose to link the smoothing priors of one time scale and present an extended multivariate approach based on correlated smoothing priors. This involves a Kronecker product precision matrix composed of the inverse of a uniform correlation matrix and the precision matrix of the univariate RW2, which results in a multivariate random walk. In the context of time series analysis, the use of multivariate random walks plays a fundamental role in multivariate modelling (Harvey, 1990). The multivariate random walk is an example of an improper correlated Gaussian Markov random field (GMRF) model. Correlated GMRF models with conditional autoregressive (CAR) structure are sometimes called multivariate CAR (MCAR) models, see for example Gelfand and Vounatsou (2003) or Carlin and Banerjee (2003). Proper multivariate GMRF models have been introduced by Mardia (1988). Greco and Trivisano (2009) applied MCAR models to handle general forms of spatial dependence occurring in multivariate spatial modelling of area data. Within spatial APC models Lagazio *et al.* (2003) and Schmid and Held (2004) used Kronecker product precision matrices to model different types of space-time interactions (Knorr-Held, 2000). However, as far as we know, correlated RW2s have never been used in standard multivariate APC models. In addition to correlated smoothing priors, we further propose the incorporation of correlated overdispersion. Actually the use of correlated overdispersion parameters is similar in spirit to seemingly unrelated regressions, where single regression equations are linked by means of correlated error terms (Harvey, 1990).

Fully Bayesian inference is conducted by either Markov chain Monte Carlo (MCMC) or integrated nested Laplace approximations (INLA) (Rue *et al.*, 2009). Model assessment is performed

by means of deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002), proper scoring rules (Gneiting and Raftery, 2007) and the log marginal likelihood.

The paper is organised as follows. In Section 2 we review multivariate APC models and describe our extended correlated approach. Then we discuss Bayesian estimation using MCMC (Section 2.1.1) and INLA (Section 2.1.2). Section 2.2 describes how model assessment is performed. Further, we present two applications in Section 3. First, we use MCMC and INLA to analyse annual data on chronic obstructive pulmonary disease (COPD) mortality in England & Wales, given in ten-year age groups and stratified by three regions (Section 3.1). We compare relative risk estimates obtained by a standard multivariate APC model with those obtained by different correlated formulations. In this case study, the precision of relative risk estimates is considerably improved by the inclusion of correlated time effects and overdispersion parameters. In the second application we analyse overall mortality rates of Scandinavian women aged $0-84$ years during the period $1900-1999$ using INLA (Section 3.2). Age groups and periods are both given in five-year intervals, resulting in 20 periods and 17 age groups. Using multivariate APC models we infer missing data of Norwegian women for 12 periods ($1900-1959$) and all age groups. The quality of the predictions is assessed with data available from the Human Mortality Database (2010). We summarise our findings in Section 4.

2 The correlated multivariate APC model

Let n_{ijr} denote the number of persons under risk in age group i ($i = 1, \ldots, I$), period j ($j = 1, \ldots, J$) and stratum r ($r = 1, \ldots, R$). We assume that the number of disease cases or deaths y_{ijr} follows a Poisson distribution with mean $n_{ijr}\lambda_{ijr}$, where in the most general formulation

$$\eta_{ijr} = \log(\lambda_{ijr}) = \mu_r + \theta_{ir} + \varphi_{jr} + \psi_{kr}.$$

Here, μ_r is the stratum-specific intercept, and θ_{ir}, φ_{jr} and ψ_{kr} are stratum-specific age, period and cohort effects, respectively. The cohort index k is a linear function of the age index i and the

period index j. If the time interval widths of age group and period are equal $k = (I-i)+j$. For the case in which the time interval widths are unequal, a slightly different definition has to be used. Suppose age group intervals are M times wider than period intervals. Then, $k = M \times (I - i) + j$ (Heuer, 1997). We apply the usual constraints, $\sum_{i=1}^{I} \theta_{ir} = \sum_{j=1}^{J} \varphi_{jr} = \sum_{k=1}^{K} \psi_{kr} = 0$ for $r = 1, \ldots, R$, to ensure identifiability of the stratum-specific intercepts. However, parameter estimates are still not identifiable because an additional constraint is necessary to obtain a specific set of parameter estimates (Holford, 2006). In contrast, second differences or projections based on the APC model are not affected by the identifiability problem and can be uniquely determined (Clayton and Schifflers, 1987; Holford, 1985).

Simpler model formulations can be obtained, for example by assuming common age effects. Then, the linear predictor is

$$\eta_{ijr} = \log(\lambda_{ijr}) = \mu_r + \theta_i + \varphi_{jr} + \psi_{kr}. \tag{2.1}$$

It should be noted that stratum-specific differences are identifiable (independent of an additional constraint), provided that not all time effects differ across strata (Riebler and Held, 2010).

Since we are in a Bayesian context, all parameters are treated as random variables and prior distributions need to be assigned. We use independent flat priors for each stratum-specific intercept μ_r. A standard multivariate APC model assigns independent smoothing priors to the age effects $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^\top$, each stratum-specific set of period effects $\boldsymbol{\varphi}_r = (\varphi_{1r}, \ldots, \varphi_{Jr})^\top$ and cohort effects $\boldsymbol{\psi}_r = (\psi_{1r}, \ldots, \psi_{Kr})^\top$, $r = 1, \ldots, R$. Consider, for example, the period effects for a specific stratum r. The RW2 is a smoothing prior based on second differences and penalises deviations from a linear trend. This improper prior can be written as:

$$\begin{aligned} f(\boldsymbol{\varphi}_r|\kappa_\varphi) &\propto \kappa_\varphi^{(J-2)/2} \exp\left(-\frac{\kappa_\varphi}{2} \sum_{j=3}^{J} ((\varphi_{jr} - \varphi_{(j-1)r}) - (\varphi_{(j-1)r} - \varphi_{(j-2)r}))^2 \right) \\ &= \kappa_\varphi^{(J-2)/2} \exp\left(-\frac{1}{2} \boldsymbol{\varphi}_r^\top \mathbf{R}_\varphi^{(2)} \boldsymbol{\varphi}_r \right) \end{aligned}$$

with precision matrix $\mathbf{R}_\varphi^{(2)}$, which depends on an unknown precision parameter κ_φ:

$$\mathbf{R}_\varphi^{(2)} = \kappa_\varphi \begin{pmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix}.$$

The RW2 is a natural choice since it is defined on the second differences which are identifiable in the likelihood (Clayton and Schifflers, 1987). Here, we assume that κ_φ does not depend on stratum r.

We propose the use of correlated smoothing priors for stratum-specific time effects, so that a potential dependence between strata can be captured. Let $\mathbf{C} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$ denote a uniform correlation matrix, where ρ is the unknown correlation parameter, $\mathbf{I}$ the identity matrix and $\mathbf{J}$ a matrix of ones. The random walks of the stratum-specific period effects $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_R$, for example, can be correlated using the stacked vector $\tilde{\boldsymbol{\varphi}} = (\boldsymbol{\varphi}_1^\top, \ldots, \boldsymbol{\varphi}_R^\top)^\top$:

$$\begin{aligned} f(\tilde{\boldsymbol{\varphi}}|\mathbf{C}_\varphi, \kappa_\varphi) &\propto (|\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}_\varphi^{(2)}|^\star)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\varphi}}^\top \{\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}_\varphi^{(2)}\}\tilde{\boldsymbol{\varphi}}\right) \\ &= |\mathbf{C}_\varphi^{-1}|^{\frac{J-2}{2}} \cdot (|\mathbf{R}_\varphi^{(2)}|^\star)^{\frac{R}{2}} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\varphi}}^\top \{\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}_\varphi^{(2)}\}\tilde{\boldsymbol{\varphi}}\right) \\ &\propto \kappa_\varphi^{\frac{R(J-2)}{2}} |\mathbf{C}_\varphi^{-1}|^{\frac{J-2}{2}} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\varphi}}^\top \{\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}_\varphi^{(2)}\}\tilde{\boldsymbol{\varphi}}\right). \end{aligned}$$

Here, $\otimes$ denotes the Kronecker product and $|\cdot|^\star$ the generalised determinant defined as the product of all non-zero eigenvalues. The determinant $|\mathbf{C}_\varphi^{-1}|$ is $[(1+(R-1)\rho_\varphi)(1-\rho_\varphi)^{R-1}]^{-1}$, see the proof in Appendix A. This formulation corresponds to a multivariate RW2 with corre-

lated increments and is an example for an improper (intrinsic) correlated GMRF (Gelfand and Vounatsou, 2003; Rue and Held, 2005).

To adjust for unobserved heterogeneity, we introduce further stratum-specific variables z_{ijr} into the linear predictor (2.1). Typically, these overdispersion parameters are assumed to be independent Gaussian variables with mean zero and unknown variance (Besag *et al.*, 1995). We propose correlated overdispersion parameters and set $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijR})^\top \sim \mathcal{N}(0, \kappa_z^{-1}\mathbf{C}_z)$ for all i and j, where κ_z denotes the precision of the overdispersion.

All of the up to eight hyperparameters (four precisions and up to four correlations) are treated as unknown. Suitable gamma-hyperpriors are assigned to the precisions. As in Knorr-Held and Rainer (2001), we use $\mathrm{Ga}(1, 0.000\,05)$ for the precisions of the time (age, period and cohort) effects and $\mathrm{Ga}(1, 0.005)$ for the precision of the overdispersion.

Correlation parameters ρ are reparameterised using the general Fisher's z-transformation (Fisher, 1958, page 219):

$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} \qquad\qquad \rho^\star = \log\left(\frac{1 + \rho \cdot (R-1)}{1 - \rho}\right), \tag{2.2}$$

where $\rho^\star$ can take any real value. It should be noted, that this transformation ensures that ρ only takes values within the interval $(-1/(R-1), 1)$, so that $\mathbf{C}$ is positive definite without imposing an additional constraint. Using $R = 2$ in (2.2) we obtain:

$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + 1} \qquad\qquad \rho^\star = \log\left(\frac{1 + \rho}{1 - \rho}\right),$$

which is frequently used for constructing confidence intervals for ρ (Konishi, 1985). Fisher's z-transformation is a variance stabilising transformation. In a Bayesian context this transformation is of particular interest since it is linked to Jeffreys' prior. Indeed, the derivative of a variance stabilising transformation corresponds to Jeffreys' prior for the original parameter (Lehmann, 1999, pages 491-492). For example, for $R = 2$, Jeffreys' prior is $\pi(\rho) \propto 1/(1-\rho^2)$, the derivative of $\log\left(\frac{1+\rho}{1-\rho}\right)$ (Lindley, 1965, pages 215-220).

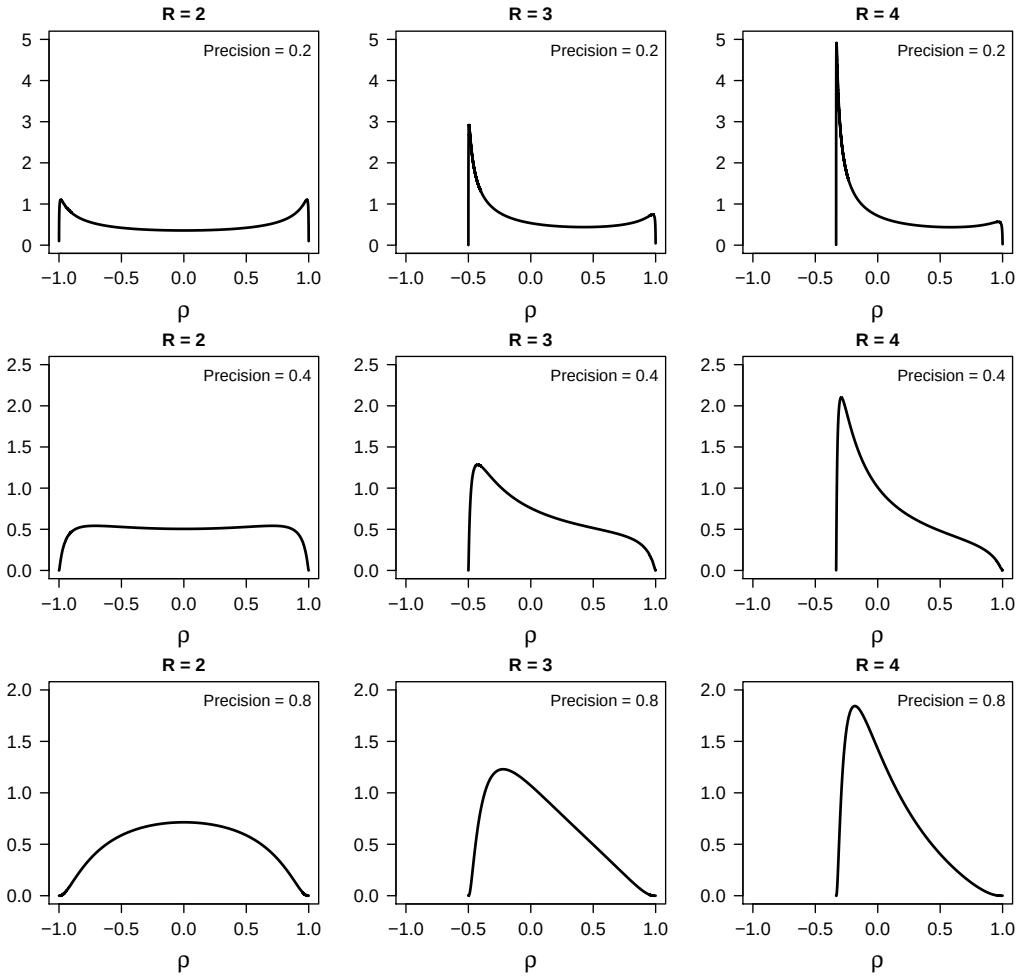

Figure 1: Prior distribution for correlation parameters ρ derived from a zero-mean Gaussian distribution for $\rho^\star$ with three different values for the precision $\kappa_{\rho^\star}$ (top to bottom: 0.2, 0.4, 0.8) and three different numbers of strata R (left to right: $R = 2$, $R = 3$, $R = 4$).

We assign a normal prior with mean zero and fixed precision $\kappa_{\rho^\star}$ to $\rho^\star$. Thus, the prior probability that ρ is larger than zero is, independent of R, equal to 0.5. Figure 1 shows for three different values of $\kappa_{\rho^\star}$ and three different values of R the resulting prior for ρ. For $R = 2$ strata, setting $\kappa_{\rho^\star}$ to 0.2 corresponds to a U-shaped prior, $\kappa_{\rho^\star} = 0.4$ to a roughly uniform prior and $\kappa_{\rho^\star} = 0.8$ to a bump-shaped prior for ρ, compare the first column of Figure 1. Note that $\kappa_{\rho^\star} = 0$ corresponds to the improper Jeffreys' prior. For a larger number of strata, the left boundary for the correlation is shifted towards zero, resulting in an asymmetric prior distribution for ρ, since

half of the total density is distributed to a smaller interval, $(-1/(R-1), 0)$. We use $\kappa_{\rho^\star} = 0.2$, so that sufficient probability mass is assigned to the boundary values as well, making extreme posterior correlation estimates possible.

2.1 Implementation

Bayesian inference for the presented models is not straightforward, since the posterior distribution is not analytically available. The common tool of choice is MCMC sampling. Recently, INLA has been presented as a new Bayesian deterministic approach. INLA replaces time-consuming MCMC sampling with fast and exact approximations to the posterior marginal distributions (Rue *et al.*, 2009). We implemented correlated multivariate APC models based on a uniform correlation structure using both MCMC and INLA. In the following we briefly present both methodologies.

2.1.1 Analysis with MCMC

Algorithmic routines based on MCMC are implemented in the low-level programming language `C` using the `GMRFLib` library (Rue and Held, 2005). Following Besag *et al.* (1995) we reparameterise the model from z_{ijr} to η_{ijr} to obtain multivariate normal full conditional distributions for the stratum-specific intercepts $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_r)^\top$ and all sets of time effects. Block updating allows the proper incorporation of the sum-to-zero constraints for the time effects. It is also possible to omit the sum-to-zero constraint for stratum-specific effects of one time scale, for example $\tilde{\boldsymbol{\varphi}}$, and simultaneously remove the stratum-specific intercepts $\boldsymbol{\mu}$ from the algorithm. For the precisions also Gibbs sampling is used. The vector $\boldsymbol{\eta}_{ij} = (\eta_{ij1}, \ldots, \eta_{ijR})^\top$ has a non-standard distribution. It is updated using multivariate Metropolis-Hastings steps with a GMRF proposal distribution based on a second-order Taylor approximation of the log likelihood (Rue and Held, 2005, Section 4.4). For the correlation parameters Metropolis-Hastings updates based on a random walk proposal are used, such that acceptance rates around 40% are achieved.

2.1.2 Analysis with INLA

A promising alternative to MCMC in the class of latent Gaussian random field models is INLA (Rue *et al.*, 2009). A detailed description of INLA and a comparison to MCMC results can be found in Rue *et al.* (2009). During a research visit of the first author in the group of Håvard Rue, correlated GMRF models were integrated into INLA enabling the analysis of correlated multivariate APC models based on a uniform correlation structure and using the general Fisher's z-transformation. However, the implementation of correlated GMRF models in INLA is not only useful in the context of multivariate APC models, but can also be used in more general structured additive regression models to correlate a wide range of GMRF models, e.g. seasonal models or models with a user-defined structure matrix. The methodology is integrated in the `R`-package `INLA` (see `www.r-inla.org`). Here, we use the `INLA` package built on 16.05.2010.

2.2 Model choice

In a Bayesian framework, frequently DIC is proposed for model comparison (Spiegelhalter *et al.*, 2002). The DIC is the sum of the posterior mean of the deviance $\bar{D}$ and the effective number of parameters p_D, that penalises increasing model complexity. Smaller values of $\bar{D}$ indicate a better model fit, but the value decreases with an increasing number of parameters. Thus, the model with the lowest DIC values provides the best trade-off between model fit and model complexity. However, in hierarchical models with many random effects, DIC has been criticised for underpenalising complex models (Plummer, 2008). Cross-validatory checks might be more appropriate, but full cross-validation is very time-consuming using MCMC.

Hence, we use approximate cross-validatory predictive checks proposed by Marshall and Spiegelhalter (2003) to calculate sound model choice criteria based on proper scoring rules in MCMC (Gneiting and Raftery, 2007). Riebler and Held (2010) used this approach in the context of standard multivariate APC models and calculated the mean ranked probability score $\overline{\text{RPS}}$ and the mean Dawid-Sebastiani score $\overline{\text{DSS}}$. To account for the correlation potentially present between multiple strata and captured by using correlated smoothing priors and correlated overdispersion

parameters we use the multivariate analogues of these measures (Gneiting and Raftery, 2007; Gneiting *et al.*, 2008). We denote these generalised forms as $\overline{\text{MDSS}}$ and $\overline{\text{MRPS}}$. Note that the $\overline{\text{MRPS}}$ corresponds to the mean energy score $\overline{\text{ES}}_{\beta=1,m=R}$ described in Gneiting and Raftery (2007, page 367).

Within MCMC we use approximate leave-one-block-out cross-validation based on replicating the vector $\boldsymbol{\eta}_{ij}$ and subsequently the observation vector $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijR})^\top$ at each iteration. In doing so we do not only lessen the influence of one observation data point y_{ijr}, but of a whole observation block $\boldsymbol{y}_{ij}$. The replicated data points can now be used to calculate the $\overline{\text{MRPS}}$. Assume we have an even number N of MCMC samples. Then

$$\overline{\text{MRPS}} = \frac{1}{I \cdot J} \sum_{i,j} \left(\frac{1}{N} \sum_{n=1}^{N} \left|\left| \boldsymbol{y}^{\text{rep}}_{ij_{(n)}} - \boldsymbol{y}_{ij} \right|\right| - \frac{1}{N} \sum_{n=1}^{N/2} \left|\left| \boldsymbol{y}^{\text{rep}}_{ij_{(n)}} - \boldsymbol{y}^{\text{rep}}_{ij_{(n+N/2)}} \right|\right| \right),$$

where $||\cdot||$ denotes the Euclidean distance and $\boldsymbol{y}^{\text{rep}}_{ij_{(n)}} = (y^{\text{rep}}_{ij1_{(n)}}, \ldots, y^{\text{rep}}_{ijR_{(n)}})^\top$ the nth replicate of observation vector $\boldsymbol{y}_{ij}$. The $\overline{\text{MDSS}}$ can be calculated as

$$\overline{\text{MDSS}} = \frac{1}{I \cdot J} \sum_{i,j} \left[\left(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}^{\text{rep}}_{ij}} \right)^\top \{\boldsymbol{\Sigma}^{\text{rep}}_{ij}\}^{-1} \left(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}^{\text{rep}}_{ij}} \right) + \log \left| \boldsymbol{\Sigma}^{\text{rep}}_{ij} \right| \right],$$

where $\overline{\boldsymbol{y}^{\text{rep}}_{ij}} = (\overline{y^{\text{rep}}_{ij1}}, \ldots, \overline{y^{\text{rep}}_{ijR}})^\top$ and $\overline{y^{\text{rep}}_{ijr}}$ is the mean of the N replicated observation samples $\boldsymbol{y}^{\text{rep}}_{ijr} = (y^{\text{rep}}_{ijr_{(1)}}, \ldots, y^{\text{rep}}_{ijr_{(N)}})^\top$ of observation data point y_{ijr}. Analogously, $\boldsymbol{\Sigma}^{\text{rep}}_{ij}$ represents the empirical covariance matrix of $(\boldsymbol{y}^{\text{rep}}_{ij1}, \ldots, \boldsymbol{y}^{\text{rep}}_{ijR})^\top$.

Thus, using MCMC we use deviance summaries and $\overline{\text{MRPS}}$ and $\overline{\text{MDSS}}$ for model comparison. In INLA deviance summaries are also returned as standard output. Furthermore, the log marginal likelihood $\log(p(\boldsymbol{y}))$ is returned. Usually the marginal likelihood is difficult to use for hierarchical GMRF models in which the underlying prior distribution is improper (here because of the RW2). However, for comparing models that only differ by the inclusion of correlation between separate priors, but have the same underlying latent structure, e.g. (2.1), $\log(p(\boldsymbol{y}))$ can be used for model choice (J.O. Berger, 2010, personal communication). Comparing two

models the posterior odds involves the ratio of their marginal likelihoods, which makes the key role of the marginal likelihood for Bayesian model comparison explicit (Bernardo and Smith, 1994, page 390).

3 Applications

We consider two applications. In the first application on COPD mortality of males in three different areas of England & Wales, we apply MCMC and INLA to estimate relative risks of death using a standard multivariate APC model and different correlated formulations. We use a MCMC run of 350 000 iterations, discarding the first 50 000 iterations and storing every 20th sample thereafter, resulting in 15 000 samples. We have routinely examined convergence and mixing diagnostics. All analyses, except of two MCMC analyses, were run on a laptop with an Intel(R) Core(TM) 2 Duo T7200 processor 2.00 GHz. The remaining two MCMC analyses were run on a server with an Intel(R) Xeon(R) X3460 processor 2.80 GHz.
In our second application on overall mortality of females in Scandinavia, we use INLA to project mortality rates of Norwegian women into the past using the existing mortality rates of Danish and Swedish women. The quality of the predictions is assessed with data available from the Human Mortality Database (2010). For both applications we apply the standard approximation settings of INLA.

3.1 COPD mortality among males in England & Wales

In this study, we analyse COPD mortality data of males in three different regions in England & Wales. COPD is a serious lung disease making it difficult to breath as a consequence of limited airflow. Known risk factors are smoking, air pollution, smog, dust, chemical fumes, etc. Data are given on an annual basis from 1950−1999 for seven age groups: 15−24, 25−34, ..., 75+ and three regions: Greater London, conurbations excluding Greater London, and nonconurbations (Hansell *et al.*, 2003; Hansell, 2004). Note that age groups (ten-year intervals) and periods (one-year intervals) are unequally spaced. Thus, there are $R = 3$ regions, $I = 7$ age groups,

$J = 50$ periods and $K = 10 \times (7 - 1) + 50 = 110$ birth cohorts.

Riebler and Held (2010) analysed these data using uncorrelated multivariate APC models. The most general model, in which all sets of parameters are assumed to be region-specific, was classified as the best model. However, since the relative risk estimates obtained by this model are not identifiable, they discussed the model with common age effects, but region-specific period and cohort effects, which was only slightly worse. Here, we compare this model (but assuming region-independent precision parameters) to three different correlated models. Either period and cohort effects, the overdispersion parameters or both, time effects and overdispersion parameters, are correlated. For all models MCMC and INLA produce virtually identical results, see Figure 2 for a comparison of precision and correlation estimates. Table 1 presents the running time and model choice criteria for all models. The running time of INLA increases with the number of hyperparameters, but is always less than the computational time of MCMC. Each model criterion prefers a different model. The $\overline{\text{MRPS}}$ prefers the uncorrelated model, while DIC classifies the model with correlated overdispersion parameters as the best. The model with both correlated time effects and overdispersion has the smallest effective number of parameters p_D, and is preferred by both $\overline{\text{MDSS}}$ and the log marginal likelihood. We regard these two scores as the most reliable since they are based on well-established theories. As can be seen from the hyperparameter estimates of the proposed model (Figure 2), the correlation estimates for ρ_φ, ρ_ψ and ρ_z are clearly different from zero confirming the classification of $\overline{\text{MDSS}}$ and $\log(p(\boldsymbol{y}))$. Hence, we regard this model as the best, and compare the relative risk estimates of this model with those obtained from the ordinary multivariate APC model. All results are presented relative to nonconurbations. The estimates of average relative risks are shown in Figure 3. The results of both models are very similar. The average relative risk of period effects shows the typical year-to-year variation with higher values in years of known air pollution events, such as the "Great Smog" in London in 1952. In the average relative risks of cohort effects different smoking behaviour may be visible. For a detailed interpretation of the relative risks we refer to Riebler and Held (2010). Because of less observations the credible intervals are getting wider

Table 1: Model choice criteria obtained by MCMC and INLA for the COPD mortality among males in England & Wales. For both approaches the running time and deviance summaries are given. In addition, the multivariate mean Dawid-Sebastiani score $\overline{\text{MDSS}}$, the multivariate mean ranked probability score $\overline{\text{MRPS}}$ and the log marginal likelihood $\log(p(\boldsymbol{y}))$ are shown. For DIC, $\overline{\text{MDSS}}$, $\overline{\text{MRPS}}$ and $\log(p(\boldsymbol{y}))$ the best value is indicated in bold. The * indicates that the analysis was run on a different machine.

		Correlation for		
	No correlation	Overdispersion	Time effects	Both
MCMC model choice				
Time	46 min 45 sec	169 min 39 sec*	53 min 36 sec	162 min 17 sec*
$\bar{D}$	1268.22	1203.94	1270.44	1252.12
p_D	424.34	419.96	415.97	394.17
DIC	1692.57	**1623.91**	1686.41	1646.29
$\overline{\text{MDSS}}$	18.87	19.23	18.96	**18.80**
$\overline{\text{MRPS}}$	**44.86**	70.65	45.64	47.51
INLA model choice				
Time	31 sec	42 sec	6 min 35 sec	68 min 3 sec
$\bar{D}$	1267.14	1203.53	1270.77	1252.09
p_D	425.30	420.13	415.66	393.97
DIC	1692.43	**1623.66**	1686.43	1646.06
$\log(p(\boldsymbol{y}))$	-4694.76	-4645.77	-4644.73	**−4632.35**

for younger birth cohorts. However, adjusting for correlation clearly improves the precision of the relative risks estimates, in particular for younger birth cohorts.

3.2 Extrapolation of overall mortality of Norwegian females

In this application, we analyse age-specific overall mortality rates of Scandinavian women. The mortality counts are stratified by five-year age group and period intervals and given for $R = 3$ regions: Denmark, Sweden and Norway (Jacobsen *et al.*, 2004). For all countries there are $I = 17$ age groups: 0–4, 5–9, ..., 75–79, 80–84. For Danish and Swedish women data exist for $J = 20$ periods 1900–1904,..., 1995–1999 and therefore $K = 36$ birth cohorts, while for Norwegian women data exist only for eight periods 1960–1964, ..., 1995–1999 resulting in $K = 24$ birth cohorts. However, for historical reasons mortality data of Norwegian women

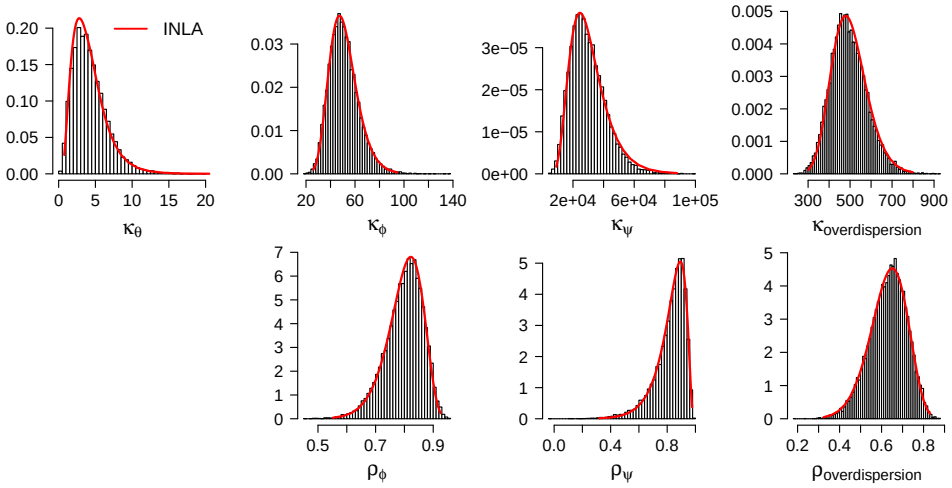

Figure 2: Approximated posterior marginals of precision and correlation parameters for the fully correlated model obtained by INLA and corresponding histograms of 15 000 MCMC samples obtained from a run with 50 000 burn-in iterations and a thinning of 20.

before 1960 would be very interesting. Figure 4 shows the mortality counts per 1 000 for all three countries stratified by age groups. The peak in mortality of both young Danish and Swedish women in the 1915–1920 period is supposed to be related to the 1918–1919 Spanish flu pandemic, which killed about 50 million people worldwide. Especially young adults showed high mortality rates. During the summer of 1918 there were strong influenza waves in Denmark, Sweden and Norway (Andreasen *et al.*, 2008; Kolte *et al.*, 2008). Thus, we expect for young Norwegian women a similar rise in mortality rates in the period 1915–1920.

Our goal is to extrapolate female mortality in Norway 1900–1959. However, extrapolation for future or past periods requires the knowledge of the corresponding population at risk for each age group. Population counts of Norwegian women for all age groups were obtained on an annual basis from Statistics Norway (2010), and were aggregated to five-year periods. We consider the model with separate age and cohort, but joint period effects, which Riebler and Held (2010) classified as the best model in an analysis of female mortality data of Denmark and Norway from 1960–1999. We compare the results of the ordinary multivariate "joint period effects" models to those obtained when using correlated time effects and overdispersion parameters.

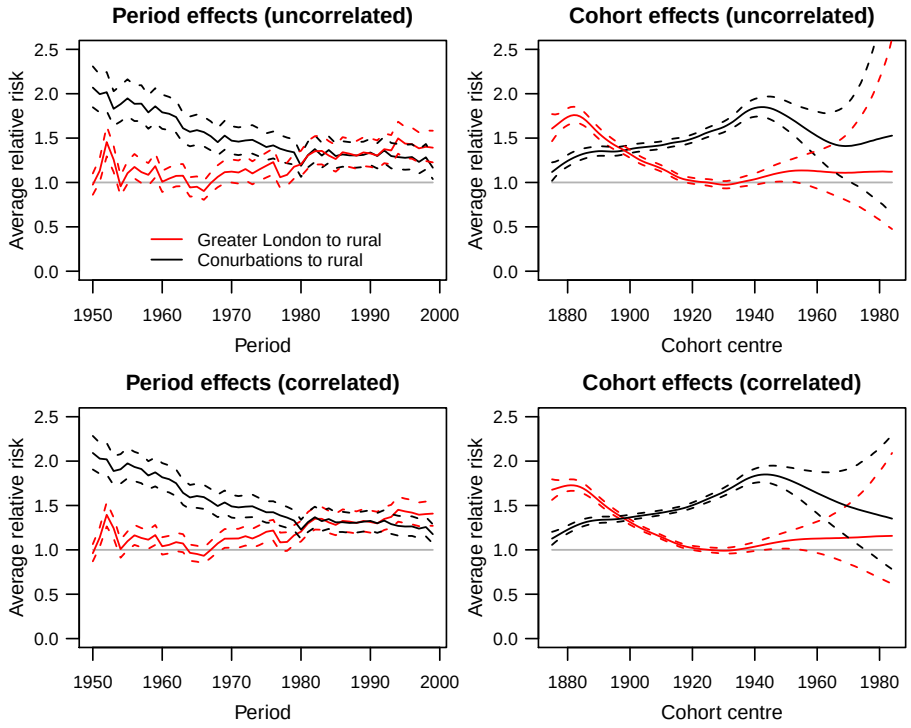

Figure 3: Average relative risk of death for Greater London and conurbations excluding Greater London compared with nonconurbations analysed by an uncorrelated multivariate "joint age effects" and a correlated "joint age effects" model assuming correlated time effects and overdispersion. Shown are the median estimates within 95% pointwise credible bands.

Table 2: Running time and model choice criteria obtained by INLA for extrapolating Norwegian mortality data. Deviance summaries and the log marginal likelihood $\log(p(\boldsymbol{y}))$ are shown. The best value for DIC and $\log(p(\boldsymbol{y}))$ is indicated in bold.

Model	Time	$\bar{D}$	p_D	DIC	$\log(p(\boldsymbol{y}))$
ordinary model	20 sec	817.66	789.19	1606.85	-6357.15
correlated model	14 min 11 sec	830.68	771.32	**1602.01**	$\mathbf{-6228.45}$

Table 2 shows deviance summaries and estimates of the log marginal likelihood for both models. The correlated "joint period effects" model is preferred by both DIC and log marginal likelihood. The posterior correlation estimates clearly indicate the dependence present between the outcomes, compare Figure 5. Figures 6–9 give observed and median predicted number of deaths per 1 000 together with 80% pointwise credible intervals for all age groups (except the last). The

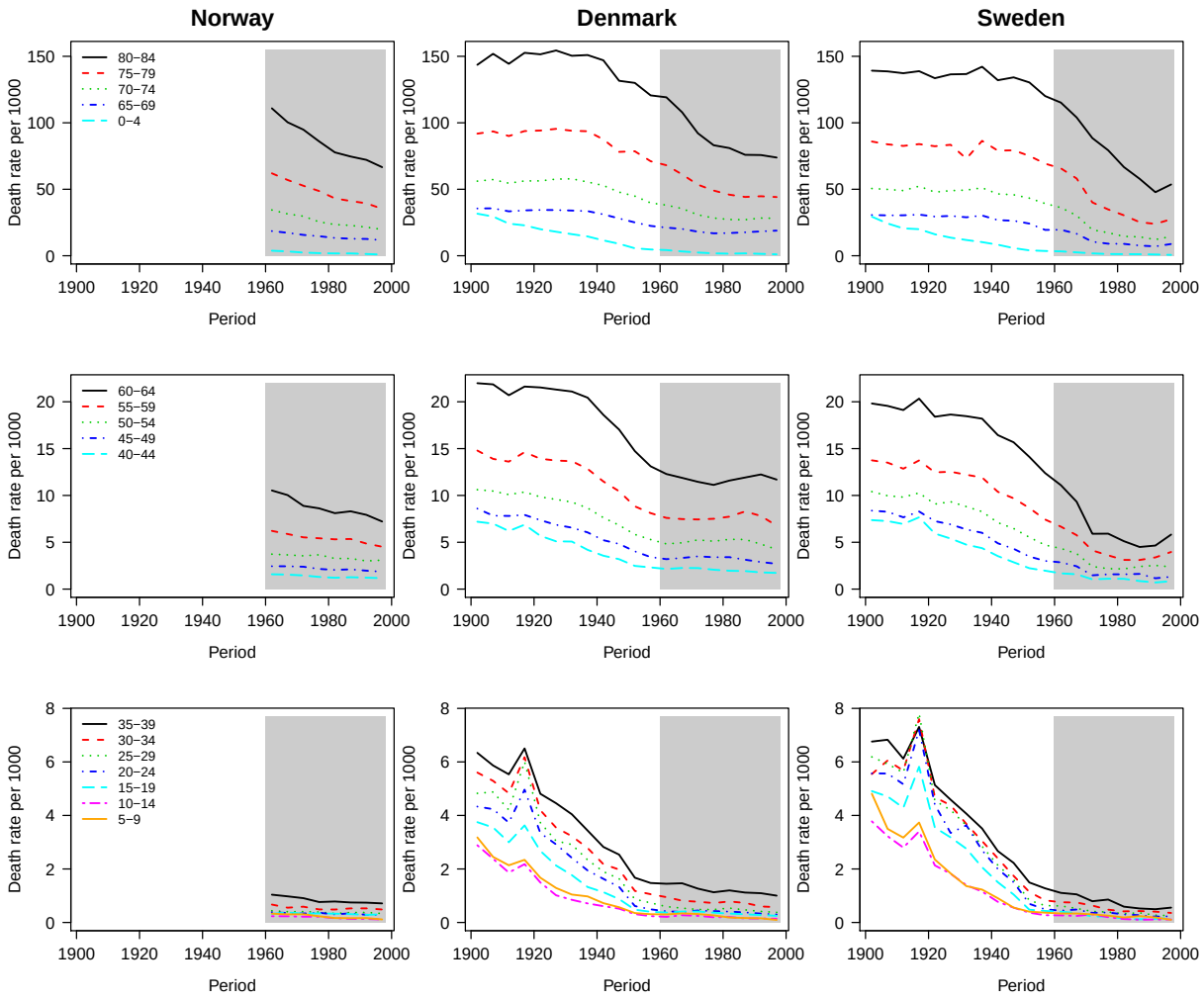

Figure 4: Mortality rates for women in Norway, Denmark and Sweden by age from 1900 to 1999. The grey shaded area shows the time range for which data for all three countries are available.

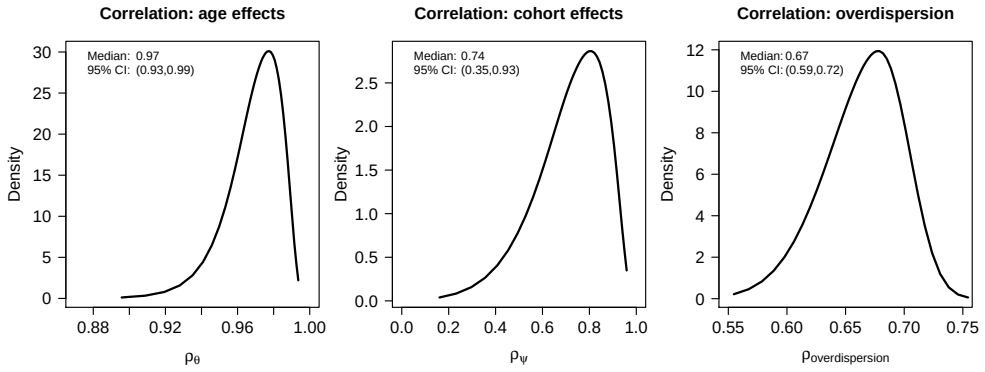

Figure 5: Posterior correlation estimates of the correlated "joint period effects" model. Approximated posterior marginals obtained by INLA are shown.

observed cases are marked by green bullets. Since INLA only returns estimates for the future linear predictors, but not for the predictive distribution of the response itself, we adjusted the predictions for the linear predictor manually for the missing Poisson variation. Otherwise, the credible intervals would be slightly narrower.

For both models, the credible intervals are getting wider as time goes into the past. Until the age of 40 years the prediction intervals overlap. However, for women older than 40 the projections of the ordinary model are clearly higher than those of the correlated model. It is interesting that the predictions of the correlated model show different temporal patterns for the different age groups, while the predictions of the uncorrelated model are almost straight lines. If the Spanish flu were captured in the period effects, the projections of the ordinary model would also show a peak for the period 1915–1920, because the period effects are assumed to be common. However, since not all age groups were affected by the influenza pandemic, the peak is not captured by the period effects, but by the overdispersion parameters. By using correlated overdispersion parameters the influenza peak is captured in the predictions for young Norwegian women obtained by the correlated model.

After we had projected the mortality rates of Norwegian women, we found in the Human Mortality Database (2010) (`www.mortality.org`) mortality counts for Norwegian women stratified

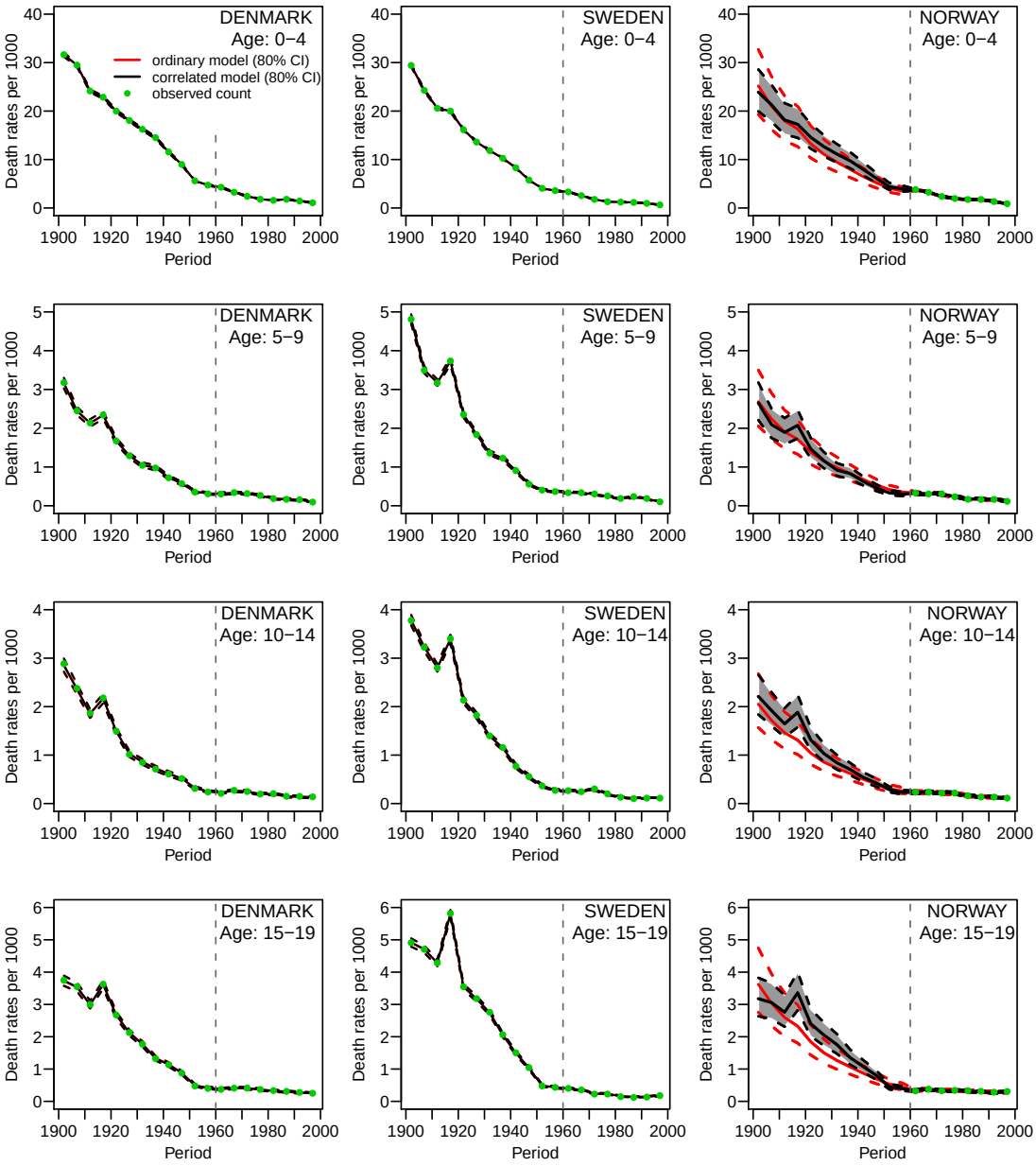

Figure 6: Observed and median predicted mortality rates within 80% credible regions in age groups 0–4, 5–9, 10–14 and 15–19 obtained by a correlated and uncorrelated multivariate APC model. Left panel: Denmark. Middle panel: Sweden. Right panel: Norway.

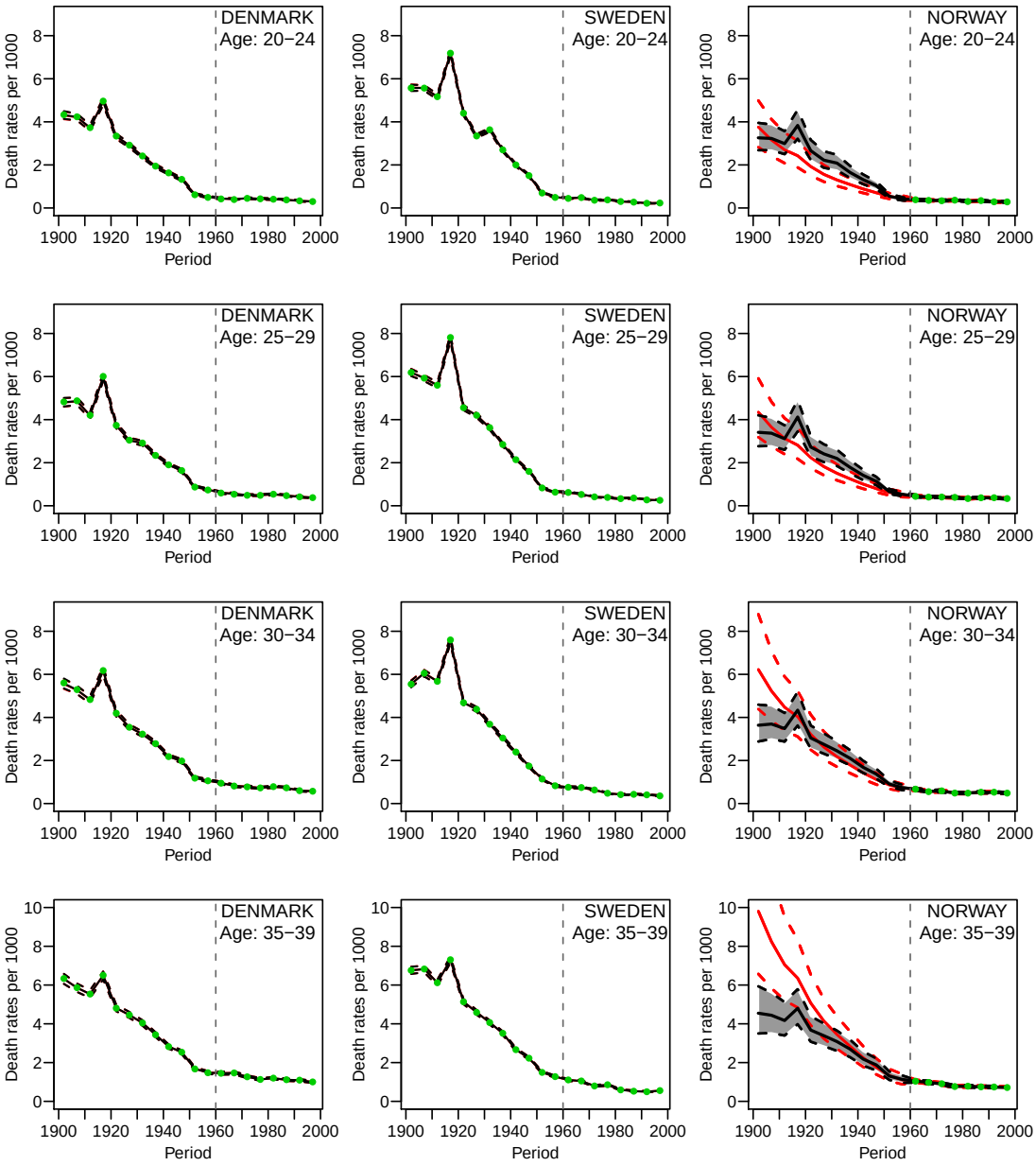

Figure 7: Observed and median predicted mortality rates within 80% credible regions in age groups 20–24, 25–29, 30–34 and 35–39 obtained by a correlated and uncorrelated multivariate APC model. Left panel: Denmark. Middle panel: Sweden. Right panel: Norway.

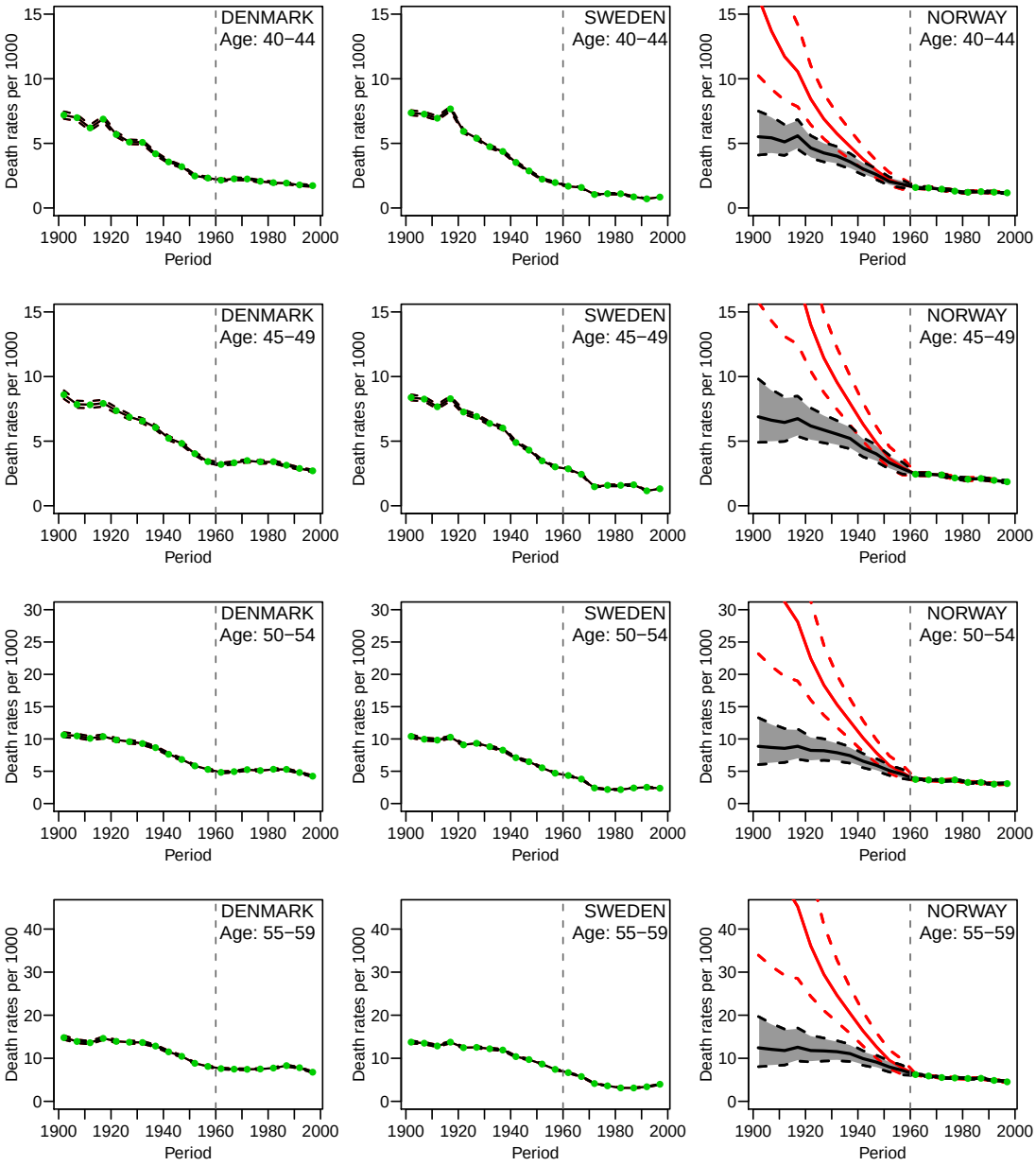

Figure 8: Observed and median predicted mortality rates within 80% credible regions in age groups 40–44, 45–49, 50–54 and 55–59 obtained by a correlated and uncorrelated multivariate APC model. Left panel: Denmark. Middle panel: Sweden. Right panel: Norway.

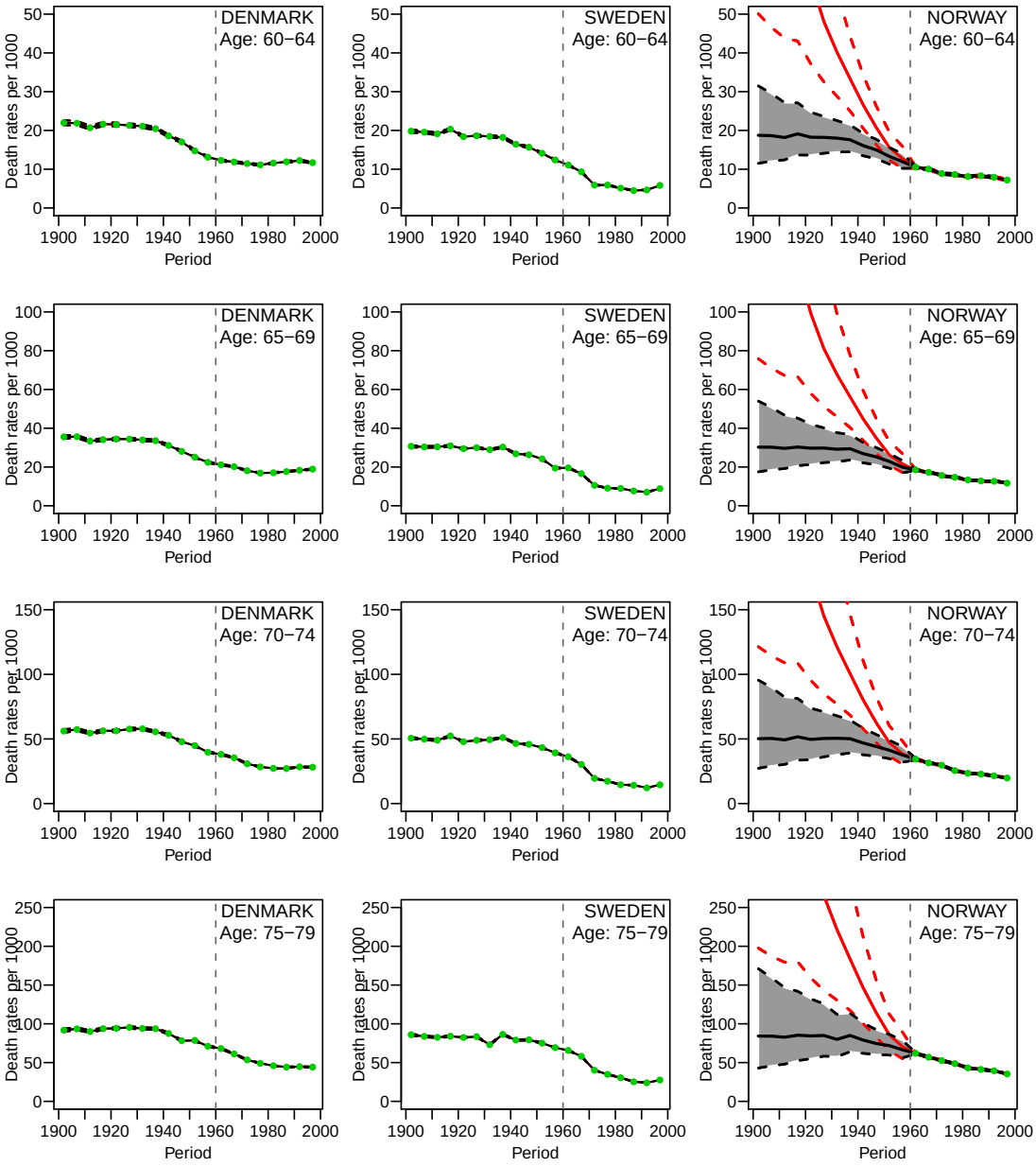

Figure 9: Observed and median predicted mortality rates within 80% credible regions in age groups 60–64, 65–69, 70–74 and 75–79 obtained by a correlated and uncorrelated multivariate APC model. Left panel: Denmark. Middle panel: Sweden. Right panel: Norway.

by five-year age group and period intervals since 1864. Thus, we assessed the quality of the predictions using these data. However, it should be noted, that although the Human Mortality Database provides the mortality rates for Danish and Norwegian women as well, not all rates are available for the same time period. Hence, nevertheless a projection of the rates for one country given the two others would have been interesting.

Figures 10 and 11 compare the predictions of the correlated and uncorrelated model with the Human Mortality Database data. For age groups below 40, both models tend to underestimate the overall mortality rate. However, the temporal patterns and especially the peak of the Spanish flu are very well captured by the correlated model. For older age groups the projections of the correlated model coincide very well with the Human Mortality Database data. In contrast, the projections of the ordinary model are too high.

4 Discussion

In this paper, we proposed the use of correlated smoothing priors and correlated overdispersion parameters for multivariate APC models. The former involves a Kronecker product precision structure for the outcome-specific time effects, i.e. age, period and/or cohort effects. We implemented correlated multivariate APC models based on a uniform correlation structure in MCMC and INLA. In the first application we analysed COPD mortality among males in England & Wales using MCMC and INLA, and compared the results of an ordinary multivariate APC model with those obtained from different correlated model formulations. A comparison of MCMC and INLA showed virtually identical results. However, the computation time of INLA was always smaller. By means of multivariate scoring rules, the log marginal likelihood and the inspection of posterior correlation estimates, the model formulation with both correlated overdispersion and correlated time effects was classified as the best. As shown in the relative risk estimates, the correlated model structure considerably improved the precision of the relative risk estimates compared to the ordinary model.

In a second application on overall mortality of Scandinavian women, we used INLA to ex-

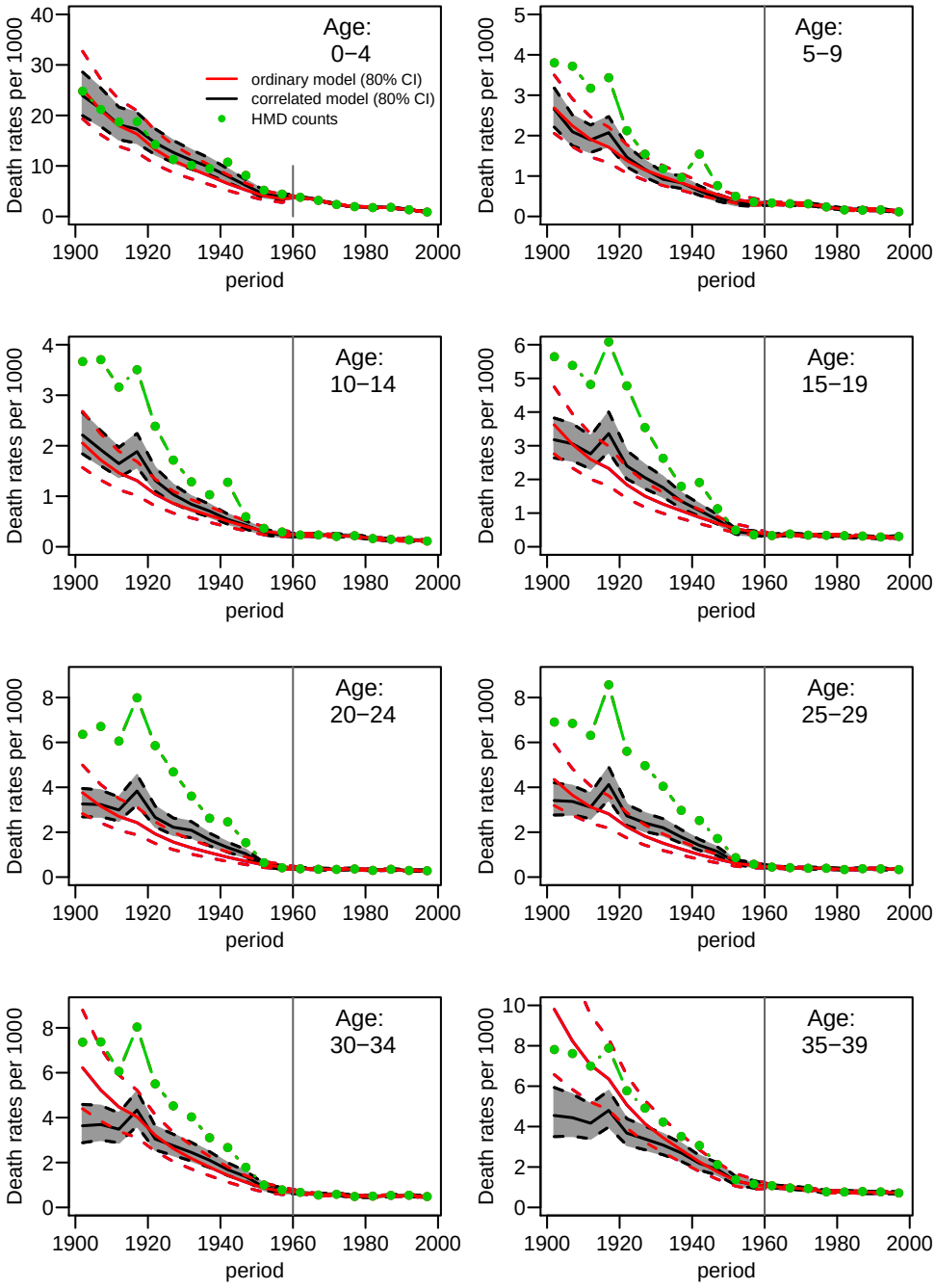

Figure 10: Median predicted mortality rates of Norwegian women within 80% credible regions for age groups 0–4, ..., 35–39 obtained by a correlated and uncorrelated multivariate APC model. In addition data available from the Human Mortality Database (2010) (HMD) are shown.

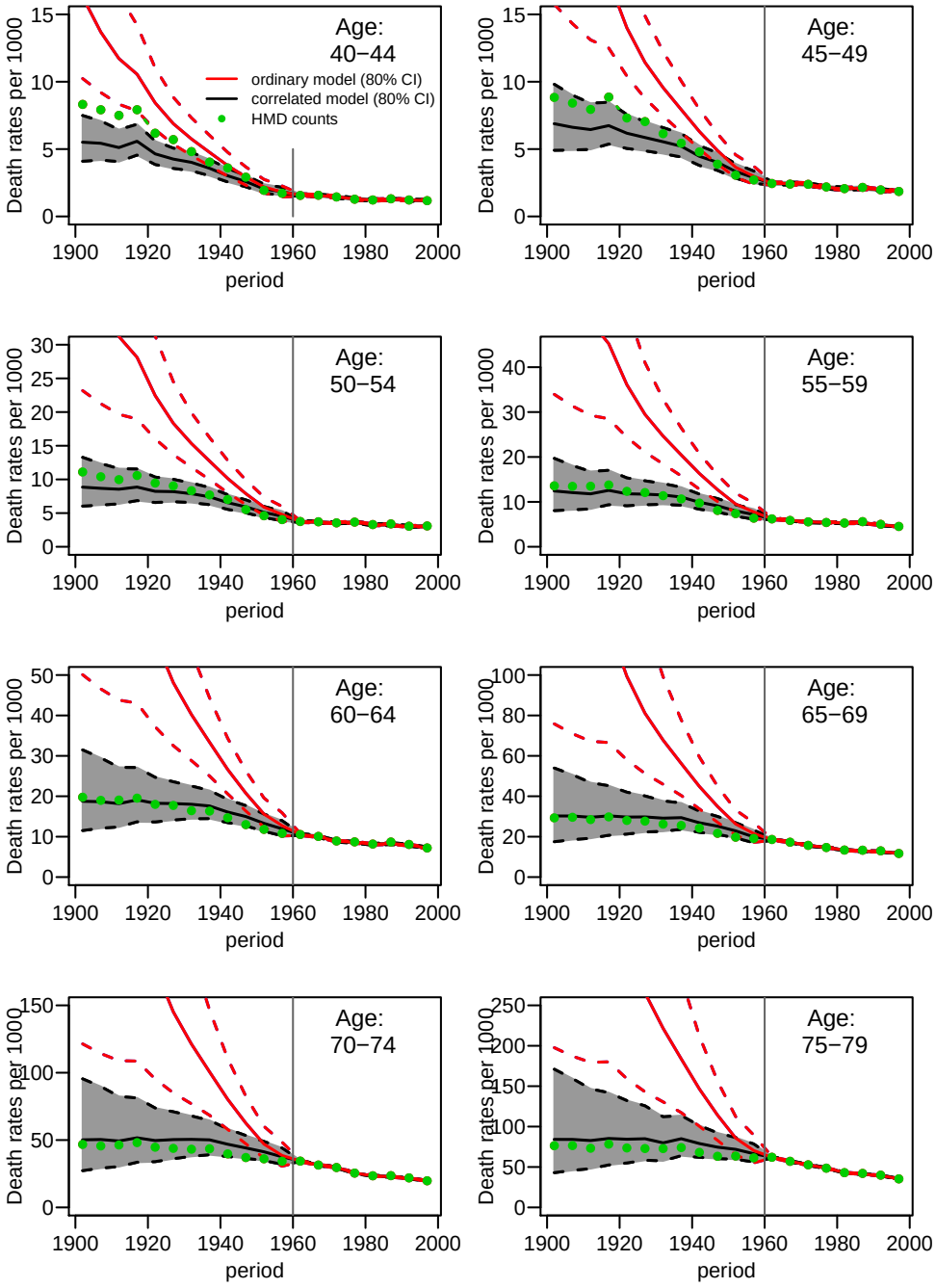

Figure 11: Median predicted mortality rates of Norwegian women within 80% credible regions for age groups 40–44, ..., 75–79 obtained by a correlated and uncorrelated multivariate APC model. In addition data available from the Human Mortality Database (2010) (HMD) are shown.

trapolate mortality rates of Norwegian women into the past by means of a joint analysis with complete age-specific rates of Danish and Swedish women. Data were available in five-year time intervals for age group and period with a total of 17 age groups and 20 (1900–1999) periods for Denmark and Sweden, and 8 periods (1960–1999) for Norway. We compared the projections of Norwegian mortality rates for 1900–1959 obtained by an ordinary uncorrelated model with those from a correlated formulation and showed that the correlated structure is useful to forecast missing data. The joint analysis of all datasets borrowed strength from shared period effects and, in the case of the correlated formulation, especially from the correlated overdispersion parameters. Thus, time patterns such as the Spanish flu were well captured in the predictions of the correlated model. Note that MCMC can also be used for this application. However, applying MCMC, problems in the mixing and convergence of single parameters, especially of the Norwegian intercept, occurred. A longer burn-in period and better initial values may help in this context.

Longterm predictions of mortality or disease rates into the future are difficult to made using the proposed approach, since usually there are no data from comparable time-series available. For short term predictions, data for some strata may be already available while for others they are still missing. Here, the correlation approach will be useful to forecast the missing units.

Another interesting field of application is similar in spirit to the inference on collapsed margins, proposed by Byers and Besag (2000). In the context of collapsed margins, complete data are available on several risk factors, but a subsequent analysis indicates that information on an additional variable is relevant. For this variable the numbers of persons at risk are available but not the numbers of cases. Byers and Besag (2000) propose a Bayesian method to estimate the effect of the variable. In the context of multivariate APC models a similar problem may occur. Here, sets of age-specific rates are available for a specific time period. However, it might be the case that these multiple data sets are only available since a specific date in time, while, before this date, data were only available for the conjunction of outcomes. A typical example could be Korea which was formerly unified, but is now divided into two states. If age-specific data on

the number of risks for the two states were available for the time when they were unified, it may be possible to project mortality rates for both states separately into the past, by exploiting the correlation present since they are divided. That means the observations for the conjunctions of both states could be separated. However, further investigations are required to explore the applicability.

The use of a Kronecker product structure is very exciting, as different correlation structures can easily be combined with different precision matrices. Based on the uniform correlation structure INLA can, by now, correlate a wide range of other GMRF models as components of more general additive regression models. Examples are: non-parametric seasonal models, continuous-time random walks or models with a user specified precision matrix. However, the uniform correlation structure is rather restrictive and may only be plausible for a few outcomes. Future work encompasses the integration of other correlation structures, for example depending on the distance between units, so that the approach can be extended to the space-time context, for example.

Acknowledgements

Financial support by the Swiss National Science Foundation and the Research Council of Norway is gratefully acknowledged.

A Uniform correlation structure

Let $\mathbf{C}$ be a $R \times R$ correlation matrix with uniform correlation structure, so that $\mathbf{C} = (1-\rho)\mathbf{I} + \rho\mathbf{J}$:

$$\mathbf{C} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

where ρ is the correlation parameter, $\mathbf{I}$ denotes the $R \times R$ identity matrix and $\mathbf{J}$ a $R \times R$ matrix of ones. Then the inverse $\mathbf{C}^{-1}$ is given by:

$$\mathbf{C}^{-1} = \begin{pmatrix} a & b & \cdots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{pmatrix} \quad \text{with} \quad \begin{aligned} a &= -\tfrac{(R-2)\cdot\rho+1}{(\rho-1)\{(R-1)\cdot\rho+1\}} \\ b &= \tfrac{\rho}{(\rho-1)\{(R-1)\cdot\rho+1\}}. \end{aligned}$$

Proof. If $\mathbf{C}^{-1}\mathbf{C} = \mathbf{I}$ then $\mathbf{C}^{-1}$ is the inverse of $\mathbf{C}$. For the diagonal elements of $\mathbf{C}^{-1}\mathbf{C}$ it follows:

$$\begin{aligned} (\mathbf{C}^{-1}\mathbf{C})_{(i,i)} &= a + (R-1)\cdot b \cdot \rho \\ &= \frac{-(R-2)\cdot\rho - 1 + (R-1)\cdot\rho^2}{(\rho-1)\{(R-1)\cdot\rho+1\}} \\ &= \frac{-R\rho + 2\rho - 1 + R\rho^2 - \rho^2}{R\rho^2 - \rho^2 - R\rho + \rho + \rho - 1} = 1 \end{aligned}$$

for all $i = 1, \ldots, R$. For the non-diagonal elements, i.e. $i \neq j$, we get:

$$\begin{aligned} (\mathbf{C}^{-1}\mathbf{C})_{(i,j)} &= a\cdot\rho + b + (R-2)\cdot b\cdot\rho \\ &= \frac{\{-(R-2)\cdot\rho - 1\}\rho + \rho + (R-2)\cdot\rho^2}{(\rho-1)\{(R-1)\cdot\rho+1\}} \\ &= \frac{-R\rho^2 + 2\rho^2 - \rho + \rho + R\rho^2 - 2\rho^2}{(\rho-1)\{(R-1)\cdot\rho+1\}} = 0 \end{aligned}$$

□

The determinant $|\mathbf{C}^{-1}|$ is given by:

$$|\mathbf{C}^{-1}| = |\mathbf{C}|^{-1} = [(1 + (R-1)\rho)(1-\rho)^{R-1}]^{-1}$$

Proof. We show that $|\mathbf{C}| = (1 + (R-1)\rho)(1-\rho)^{R-1}$, as the inverse case follows immediately. Remember that $|\mathbf{C}| = |\mathbf{I} - \rho\mathbf{I} + \rho\mathbf{J}|$. The identity matrix has R times the eigenvalue 1. The matrix $\mathbf{J}$ has once the eigenvalue R and $R-1$ times the eigenvalue 0. Since both matrices ($\mathbf{I}$ and $\mathbf{J}$) share the same eigenvectors, the eigenvalues for $\mathbf{C}$ are $(1 - \rho + \rho \cdot R)$ and $(1-\rho)$ with multiplicity $R-1$, so that the determinant of $\mathbf{C}$, the product of the eigenvalues, is:

$$|\mathbf{C}| = (1 - \rho + \rho \cdot R)(1-\rho)^{R-1} = (1 + (R-1)\rho)(1-\rho)^{R-1}.$$

□

References

Andreasen, V., Viboud, C. and Simonsen, L. (2008). Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: implications for pandemic control strategies, *Journal of Infectious Diseases* **197**(2): 270–278.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, John Wiley & Sons, Ltd, Chichester.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Baysesian computation of stochastic-systems, *Statistical Science* **10**(1): 3–41.

Bray, I. (2002). Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality, *Journal of the Royal Statistical Society - Series C* **51**: 151–164.

Byers, S. and Besag, J. (2000). Inference on a collapsed margin in disease mapping, *Statistics in Medicine* **19**: 2243–2249.

Carlin, B. P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data, *in* J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman and A. F. M. Smith (eds), *Bayesian Statistics 7*, Oxford University Press, Oxord, pp. 45–63.

Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models, *Statistics in Medicine* **6**(4): 469–481.

Fisher, R. A. (1958). *Statistical methods for research workers*, 13th ed.(rev.) edn, Oliver & Boyd, Edinburgh.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics* **4**(1): 11–25.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* **102**(477): 359–378.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L. and Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds, *Test* **17**(2): 211–235.

Greco, F. P. and Trivisano, C. (2009). A multivariate CAR model for improving the estimation of relative risks, *Statistics in Medicine* **28**(12): 1707–1724.

Hansell, A., Knorr-Held, L., Best, N., Schmid, V. and Aylin, P. (2003). COPD mortality trends 1950-1999 in England & Wales - did the 1956 Clean Air Act make a detectable difference?, *Epidemiology* **14**(5): S55–S55.

Hansell, A. L. (2004). *The Epidemiology of Chronic Obstructive Pulmonary Disease in the*

UK: Spatial and Temporal Variations, PhD thesis, Faculty of Medicine of the University of London, Imperial College, St Mary's Campus.

Harvey, A. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*, reprinted edn, Cambridge University Press, Cambridge.

Heuer, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines, *Biometrics* **53**(1): 161–177.

Holford, T. R. (1985). An alternative approach to statistical age-period-cohort analysis, *Journal of Chronic Diseases* **38**(10): 831–836.

Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals, *Statistics in Medicine* **25**(6): 977–993.

Human Mortality Database (2010). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 30th of March 2010).

Jacobsen, R., Von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different?, *European Journal of Epidemiology* **19**(2): 117–121.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in Medicine* **19**: 2555–2567.

Knorr-Held, L. and Rainer, E. (2001). Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction, *Biostatistics* **2**(1): 109–129.

Kolte, I. V., Skinhoj, P., Keiding, N. and Lynge, E. (2008). The Spanish flu in Denmark, *Scandinavian Journal of Infectious Diseases* **40**: 538–546.

Konishi, S. (1985). Normalizing and variance stabilizing transformations for intraclass correlations, *Annals of the Institute of Statistical Mathematics* **37**: 87–94.

Lagazio, C., Biggeri, A. and Dreassi, E. (2003). Age-period-cohort models and disease mapping, *Environmetrics* **14**(5): 475–490.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*, Springer, New York.

Lindley, D. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge University Press, Cambridge.

Mardia, K. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing, *Journal of Multivariate Analysis*.

Marshall, E. C. and Spiegelhalter, D. J. (2003). Approximate cross-validatory predictive checks in disease mapping models, *Statistics in Medicine* **22**(10): 1649–1660.

Osmond, C. (1985). Using age, period and cohort models to estimate future mortality rates, *International Journal of Epidemiology* **14**(1): 124–129.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison, *Biostatistics* **9**(3): 523–539.

Riebler, A. and Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models, *Biostatistics* **11**(1): 57–69.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC Press, London.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society - Series B* **71**: 319–392.

Schmid, V. and Held, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data, *Biometrics* **60**: 1034–1042.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R. and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society - Series B* **64**: 583–616.

Statistics Norway (2010). 05839: Population, by sex and age (1846-2009). Data downloaded on 24th January 2010. Available from: `www.sbb.no/english/`.

PAPER IV

Suicide mortality in Switzerland: Gender-specific differences and the impact of family integration

Andrea Riebler, Leonhard Held, Håvard Rue & Matthias Bopp

Technical Report, University of Zurich.

Suicide mortality in Switzerland: Gender-specific differences and the impact of family integration

Andrea Riebler[1]*, Leonhard Held[1], Håvard Rue[2] and Matthias Bopp[1]

[1]Institute of Social and Preventive Medicine,

University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland;

[2]Department of Mathematical Sciences, Norwegian University of Science and Technology,

N-7491 Trondheim, Norway;

Suicide has become one of the leading causes of death of Swiss males aged between 15 and 44 years. The age-standardised rates are about three times higher than for females. We compare suicide mortality of males and females in Switzerland and investigate gender-specific differences over the last 58 years. Multivariate age-period-cohort (APC) models are applied to jointly analyse age-specific suicide rates for Swiss men and women aged 15–79 during the period 1950–2007. This approach avoids age aggregation and explores heterogeneous time trends across age, period and birth cohort. We found strong gender-specific differences in suicide mortality. While the same risk factors may act on age and overdispersion there is no significant correlation between cohort effects. Following the well-known Durkheimian theory, we further explore whether measures of family integration can explain suicide trends of males and females. Here, APC models provide a new way to explain suicide mortality through explanatory variables. Effects of covariates related to family integration are found to be similar for males and females and exert an inverse influence on suicide risk. However, since suicide is influenced by a combination of different risk factors, a measure of social integration can only partially explain the underlying trends.

Keywords: Age-Period-Cohort model; Suicide; Switzerland; Bayesian; Family integration.

*To whom correspondence should be addressed. Email: andrea.riebler@ifspm.uzh.ch

1 Background and objectives

Age and gender are well established risk factors of suicide. In many countries especially elderly have a higher risk to commit suicide (Granizo *et al.*, 1996; Shah and De, 1998). Suicide rates of males are generally higher than those of women. Besides age- and gender-specific factors, environmental factors occurring at a specific date in time may influence suicide risk. For example, Lester (1990) examined suicide rates in Switzerland during the period when domestic gas was detoxified. This study did not only indicate a decline of suicide by means of domestic gas, but also a decline of the overall suicide rate, indicating that people did not switch to alternative methods. Thus, if one method of suicide is made less available, such as detoxification of domestic gas or strict gun control rules, a reduction in the overall suicide risk might be observed. Experiences common to a particular birth cohort, e.g. war, might also influence the suicide risk. Hence a better understanding of age, period and cohort effects might help to target effective preventive strategies. Over the last years a number of age-period-cohort (APC) analyses of suicide rates have been published, see for example Granizo *et al.* (1996); Etzersdorfer *et al.* (1996); Snowdon and Hunt (2002); Stockard and O'Brien (2002); Gunnell *et al.* (2003); Ajdacic-Gross *et al.* (2006).

Although Switzerland is an affluent country the suicide rates observed are quite high compared to other countries (Levi *et al.*, 2003). To gain more information on gender-specific differences in Switzerland Ajdacic-Gross *et al.* (2006) performed univariate APC analyses on suicide data over the last century. They found similar age and period effects, but stronger cohort effects for males than for females. However, since males and females might be subject to similar risk factors it seems justified to model them jointly suggesting the possibility of common time (age, period, cohort) effects.

In this paper we will apply multivariate APC models to capture trends in the sex ratio. Such models borrow strength from both genders for estimating common sets of time effects, for example the age effects, while the remaining parameter sets can be different across gender. The well-known identifiability problem of APC models is avoided since differences of gender-specific

time effects are identifiable and can be interpreted as log relative risk (Riebler and Held, 2010). Since social aspects are strongly associated with suicide risk, we will further investigate whether changes in variables related to family integration can explain suicide trends. Following the theory of Durkheim (Durkheim, 1897) numerous papers were published investigating the relationship between marital status and suicide mainly based on time-series analyses (Breault and Barkey, 1982; Lester, 1986; Stack, 1987, 1990a,b, 1992; Surault, 1992; Rossow, 1993; Lester, 1994). Populations with high divorce rates, for example, were found to have high suicide rates even when controlling for confounding variables such as socio-economic status (Stack, 1990a). Nowadays only few publications investigate a correlation between family integration and suicide. Stack (1990b) assumed that the association between being divorced and committing suicide has changed over the years. For example, divorce and unmarried couples are more common and more accepted. We will apply univariate APC models separately to males and females replacing the period effects by an explanatory variable on family integration assuming either a parametric or non-parametric effect. Replacing one set of effects by an explanatory variable is also a valid solution to the non-identifiability problem of APC models (Brown and Kessler, 1988). However, this is only the case if the covariate-effect does not depend on the time scale for which it is used. Otherwise a linear dependence between the three time scales remains, see the introduction of Riebler (2010, page 4) for a discussion of this topic. Further, we apply a multivariate age-cohort (AC) model replacing the period effects by a correlated non-parametric covariate effect of family integration. The rate of unemployment is included in the model formulation to account for confounding. However, as the Swiss unemployment rate is so low, variation may be quite small to affect the overall suicide trends (Stack, 1989).

2 Data and Methods

Annual age- and gender-specific suicide mortality counts and mid-year population data, 1950–2007, were obtained from the Swiss Federal Statistical Office (statistics of causes of death). Age groups are stratified by five-year intervals: 15–19, ..., 75–79, resulting in 13 age groups and

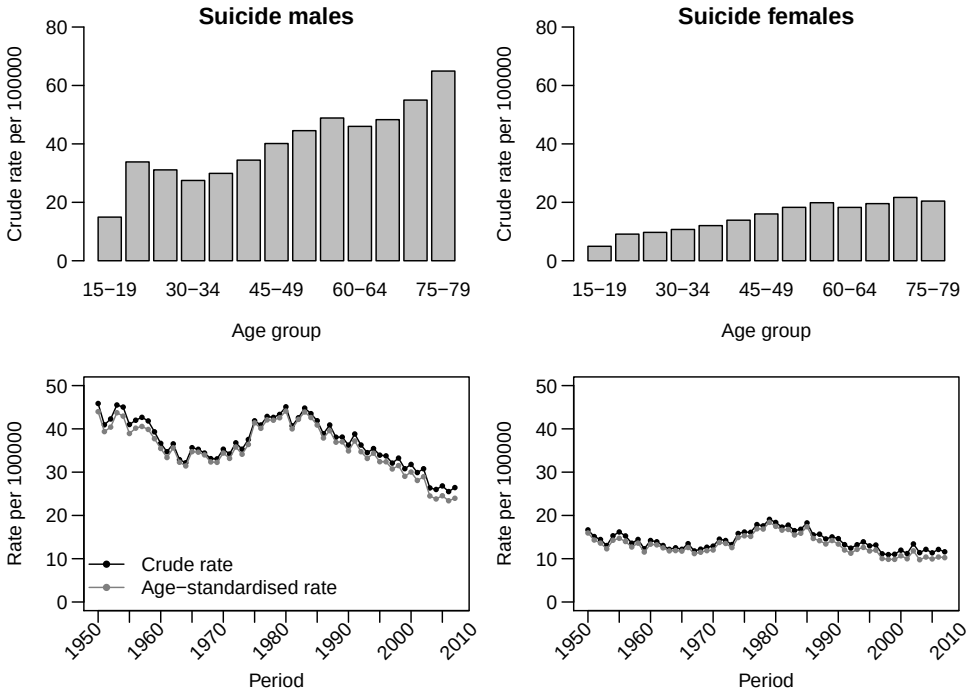

Figure 1: Suicide rates from 1950-2007. Top: Age-specific crude rates for males and females. Bottom: Crude and age-standardised rates of all periods for males and females.

58 one-year periods. We omitted data for children under the age of 15, when suicide is rare, and for adults over the age of 79 because for elderly assisted suicides have become frequent in Switzerland. Figure 1 shows age-specific crude rates and both crude and age-standardised rates per 100 000 persons of all periods for males and females. Age-standardisation was performed using the WHO world standard population (Ahmad *et al.*, 2001).

First, the data were analysed using Bayesian APC models for multiple outcomes (Riebler and Held, 2010). We assume that the number of suicides y_{ijg} of age group $i = 1, \ldots, 13$, period $j = 1, \ldots, 58$ and sex g is Poisson distributed with rate $n_{ijg} \times \lambda_{ijg}$. Here n_{ijg} is the number of persons at risk and $\log(\lambda_{ijg})$ denotes the linear predictor. Under the assumption of joint age effects, for example, differences of gender-specific period and birth cohort estimates are identifiable. The linear predictor is

$$\log(\lambda_{ijg}) = \mu_g + \theta_i + \varphi_{jg} + \psi_{kg} \tag{2.1}$$

where μ_g represents the gender-specific mean (intercept), θ_i the common age effect for age group i, φ_{jg} the effect of period j for sex g and ψ_{kg} the effect of the kth cohort for sex g. Note that the cohort index $k = 1, \ldots, K$ depends on the age index i and period index j, but also on the width of age group and period intervals. We use the definition of Heuer (1997) which results in $K = 118$ birth cohorts. For identifiability of the gender-specific intercepts, we constrain all sets of time effects to sum to zero, i.e. here $\sum_{i=1}^{I} \theta_i = 0$, $\sum_{j=1}^{J} \varphi_{jg} = 0$ and $\sum_{k=1}^{K} \psi_{kg} = 0$ for both males ($g = 1$) and females ($g = 2$). If suitable, the linear predictor (2.1) can be modified to allow for separate period but common cohort effects or vice versa. Similarly, age and cohort effects may vary across gender but the period effects may be common, for example. However, keep in mind that differences are not identifiable if all sets of effects are allowed to vary (Riebler and Held, 2010).

Since we are in a Bayesian setting we treat all parameters as random and assign prior distributions. We follow Riebler and Held (2010) and use independent flat priors for the gender-specific intercepts. For the age, period and cohort effects we expect similarities between effects adjacent in time. Thus, we choose a Gaussian prior distribution based on independent second differences for all time effects, also known as random walk of second order (Besag *et al.*, 1995). This is a natural choice since second differences of time effects are identifiable (Clayton and Schifflers, 1987). Note that variance parameters are not gender-specific. Thus, there is one variance parameter for each time scale resulting in three variance parameters in total. The variances are treated as random and suitable prior distributions are assigned.

As is well known when working with registry data there are often inconsistencies in reporting systems or changes in reporting behaviour (Brillinger, 1986). For example the coding of causes of death might change over time. To adjust for such overdispersion, i.e. unobserved heterogeneity, we introduce further gender-specific variables z_{ijg} with mean zero and unknown variance into the linear predictor (2.1) (Besag *et al.*, 1995).

To validate and compare the "joint age effects" model (2.1) with other models we use the well-known deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002). However, this criterion

has recently been criticised for models with many random effects, e.g. (2.1), because complex models tend to be under-penalised (Plummer, 2008). For this reason we additionally calculate cross-validated proper scoring rules (Gneiting and Raftery, 2007), such as the mean Dawid-Sebastiani score (mean DSS), the mean ranked probability score (mean RPS) and the log score. Both DIC and proper scoring rules are negatively oriented such that smaller values are better. Having found the best model the question arises whether it is necessary to allow for correlation between gender-specific time effects and/or gender-specific overdispersion parameters. A model with correlated overdispersion parameters is similar in spirit to a seemingly unrelated regression models (Zellner, 1962). In this class of models, there are several regression equations which are assumed to be correlated via their error terms. Regarding the time effects (age, period and cohort effects) the inclusion of a correlation would actually represent a balance between separate and joint effects. More precise relative risk estimates may be obtained. For comparing models with and without correlation structure we additionally use the log marginal likelihood. Although improper priors are used, the use of the marginal likelihood is valid because the candidate models only differ by the inclusion of correlation between the priors, for more details see Riebler *et al.* (2010).

All models are estimated using both Markov chain Monte Carlo (MCMC) techniques as described in Riebler and Held (2010) and Riebler *et al.* (2010), and integrated nested Laplace approximations (INLA) (Rue *et al.*, 2009). INLA represents a deterministic alternative to MCMC. It computes directly very accurate approximations to the posterior marginal distributions and thus avoids time-consuming sampling. The INLA-program is freely available under `www.r-inla.org` and runs on Windows, Mac and Linux. Here we use the INLA version built on 16.05.2010. DIC can be calculated within both settings. The log score and marginal likelihood are calculated using INLA, in contrast mean DSS and the mean RPS are calculated with MCMC. For comparing correlated multivariate APC models the multivariate analogues of the mean RPS and DSS score are used to capture the potential correlation present between the male and female suicide rates (Riebler *et al.*, 2010).

For the MCMC analyses without inclusion of correlation we use a run of 120 000 iterations, discarding the first 20000 iterations and storing every 20th sample thereafter, leading to a total of 5000 samples. When including correlation we use a run of 520 000 iterations omitting the first 20000 iterations and storing every 200th sample thereafter, resulting in 2500 samples. This is necessary because of the high autocorrelations when accounting for correlations between time-effects and/or overdispersion parameters.

Explanatory variables on family integration are available for all observation years (1950–2007) and obtained for the 1970 to 2007 period from The Swiss Federal Statistical Office (2009) and for the period before 1970 from Calot (1998). Since social integration is difficult to measure, there are several proposals and it is debatable which one to choose as an indicator. Breault and Barkey (1982) proposed to measure family integration by the marriage rates minus divorce rates divided by marriage rates plus divorce rates, since married people are supposed to be better integrated than single persons. A high value is related to a better integration. We will call this indicator F-index throughout the paper. The F-index is preferable to a crude measure, such as divorce rates alone as there is a high correlation between divorce rates and marriage rates. Alternatively the number of divorces per 100 marriages could be used.

To calculate the F-index we first adjusted the divorce rate for the introduction of a new divorce law in 2000 in Switzerland. The new law resulted in an extreme increase of divorces in 1999 and a strong decrease in 2000. First there were more divorce proceedings terminated in 1999 to have more time to adjust to the new legal situation in 2000. Then the introduction of the new divorce law in 2000 caused a prolongation of proceedings and thus less decisions. Thus, we substituted the divorce rates in the years 1999 and 2000 by the average of this two years. Alternatively a moving average of first or second order could be applied to the whole time-series to smooth random variations over the whole period.

As an alternative measure for family integration we consider the total marriage rate. This index represents the mean percentage of unmarried persons aged below fifty who would marry in the course of time, if they showed the same age-specific marriage behaviour as in the observation

year.

For analysing the association of family integration and suicide we start with the calculation of pairwise correlations and perform standard time-series analyses, as proposed in Stack (1989, 1990a). Then, we use INLA to estimate separate univariate APC models for males and females substituting the period effect block by a time-constant regression variable related to the rate of unemployment and an explanatory variable on family integration assuming either a parametric or non-parametric effect. Finally, we use the multivariate APC model classified as the best model without covariates and replace the period effects by a non-parametric covariate effect of the F-index assuming a correlation between the effect on males and females. Figure 2 shows the F-index, the total marriage rates for males and females, and the unemployment rate from 1950–2007. Both the F-index and also the total marriage rates strongly decrease from 1950 to 2007. Especially around the middle of the 1970s there is a big drop. Around the early 1990s a local maximum is visible while afterwards all markers are decreasing again. The unemployment rates show the typical economical patterns indicated within the figure. However, note that the rates are always below five per cent.

3 Results

We will first present the results of the multivariate APC analysis for detecting trends in the sex ratio in suicide mortality of males and females. Then we will explore whether the inclusion of explanatory variables related to family integration can explain changes in suicide rates.

3.1 Analysis of gender-specific differences

The model diagnostics for all models obtained by MCMC and INLA are shown in Table 1. The "joint period effects" model is classified by all model choice criteria as the best model, so that we assume gender-specific age and cohort effects. Including a correlation for overdispersion is clearly preferred to the model without correlation. The model with correlated overdispersion and the model with both correlated gender-specific effects and overdispersion are very similarly

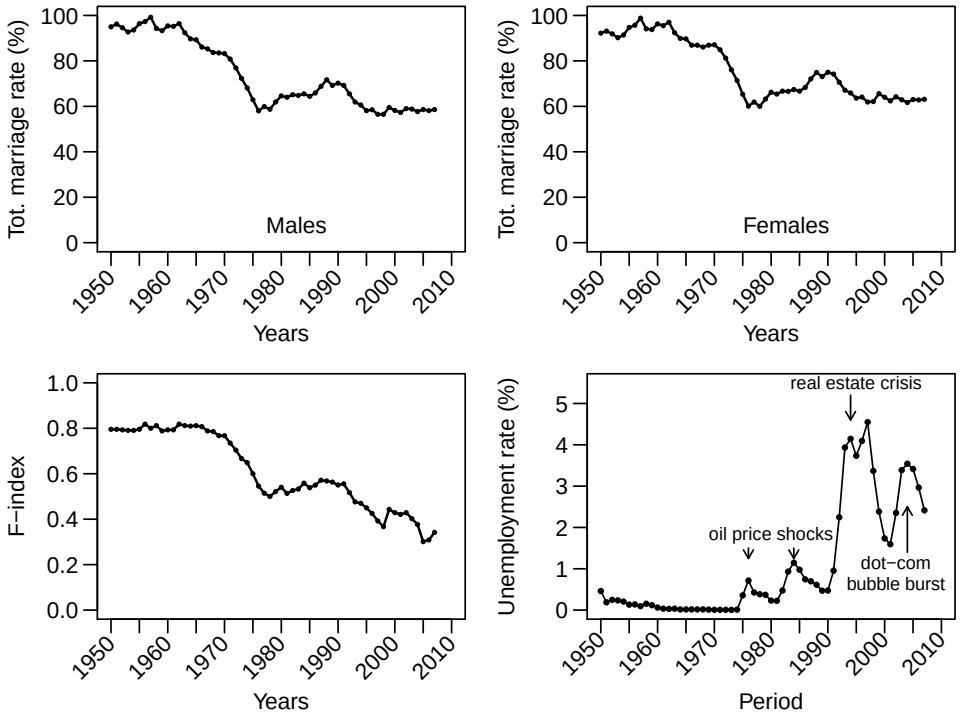

Figure 2: Covariate information from 1950–2007. Top (left to right): Total marriage rate for males and females given in per cent. Bottom (left to right): F-index measured as (marriage rate - divorce rate)/(marriage + divorce) and unemployment rate measured as unemployed/(unemployed + employee).

classified. The log-marginal likelihood slightly prefers the model with a correlation for both time scales and overdispersion, see Table 2. In this model, the correlation between overdispersion parameters was estimated as 0.56 with 95% credible interval $(0.18, 0.84)$. For the gender-specific age-effects the correlation was estimated as 0.82 $(0.50, 0.95)$ and for the cohort effects to 0.20 $(-0.61, 0.79)$. Thus, except for the cohort effects, the posterior distributions of all correlation estimates are clearly greater than zero, indicating that partly the same risk factors act on age and overdispersion for males and females.

Figure 3 shows the relative risks of suicide for males compared to females for the "joint period effects" model with both correlated age and cohort effects, and correlated overdispersion. Men have an about three times higher risk to commit suicide than women. Especially men between 16 and 25 have a high risk compared to the corresponding age group of females. While the risk

Table 1: Model choice criteria obtained from MCMC and INLA. For both approaches DIC estimates are given. In addition the mean Dawid-Sebastiani score $\overline{\text{DSS}}$ and the mean ranked probability score $\overline{\text{RPS}}$ are shown for MCMC and the log score for INLA. The column names indicate which effects (**A**ge, **P**eriod, **C**ohort) are assumed to be the same for males and females. The remaining effects are assumed to be gender-specific. The best value for each criterion is indicated in bold.

	Joint effects used for						
	A,P,C	P,C	A,C	A,P	C	P	A
MCMC model diagnostics							
Mean DSS	5.10	4.99	5.06	4.96	4.95	**4.88**	4.92
Mean RSS	4.62	4.37	4.53	4.34	4.28	**4.15**	4.21
DIC	2064.87	1971.86	2037.82	1942.87	1936.84	**1864.43**	1899.70
INLA model diagnostics							
Log Score	3.48	3.44	3.47	3.43	3.42	**3.39**	3.40
DIC	2064.68	1971.39	2037.47	1942.88	1936.74	**1864.16**	1899.29

Table 2: Model choice criteria obtained from MCMC and INLA for the "joint period effects" model without correlation, correlation between overdispersion parameters, between gender-specific age and cohort effects, between both overdispersion and gender-specific age and cohort effects. The best value for each criterion is indicated in bold.

	Correlation included for			
	-	overdispersion	time effects	both
MCMC model diagnostics				
Mean multivariate DSS	9.758	**9.752**	9.763	9.756
Mean multivariate RPS	6.653	**6.636**	6.647	6.640
DIC	1864.372	1854.804	1864.270	**1854.788**
INLA model diagnostics				
Log marginal likelihood	-5215.39	-5214.86	-5215.78	**−5214.72**
Log Score	3.390	**3.386**	3.390	3.386
DIC	1864.16	**1854.56**	1864.07	1854.60

for males between 30 and 74 is almost three times as high as in women it is increasing again for elderly men.

For cohorts born around 1870 the risk for males is about four times as high compared to females.

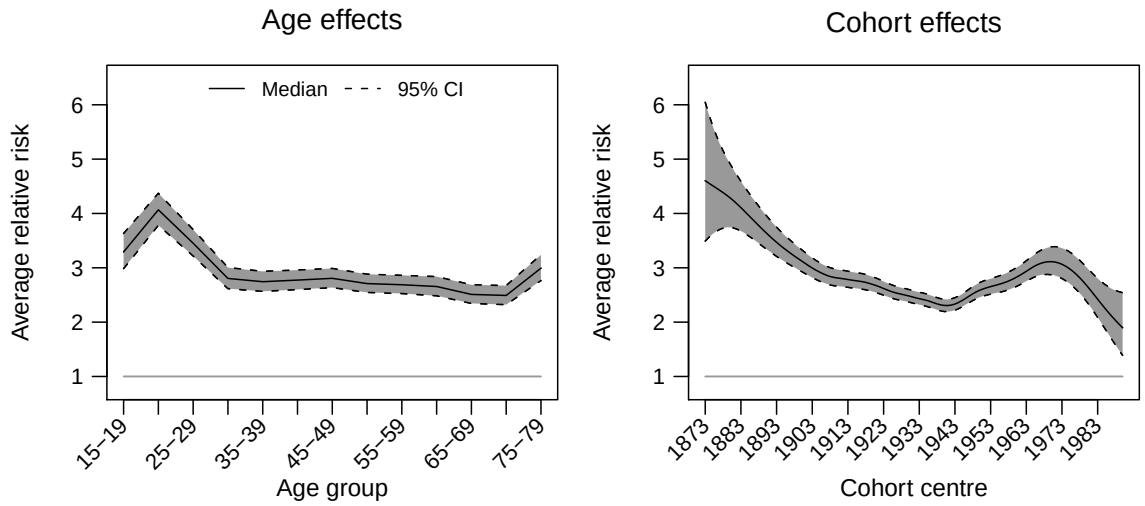

Figure 3: Relative risk of suicide for males compared to females for the "joint period effects" model assuming a correlation between males and females for the age and cohort effects, and also the overdispersion parameters.

However, with successive cohorts the average relative risk falls to being about twice as high for cohorts born around 1940. Then it increases again reaching a local maximum for cohorts born around 1970. For the youngest cohorts a decreasing trend is visible.

Posterior marginal distributions of MCMC and INLA were virtually identical for all variance parameters. For the gender-specific intercepts we found a small shift between MCMC and INLA which is probably related to the fact that the time effects (especially for the cohorts) of INLA do not exactly fulfil the sum-to-zero constraints. This may be solved by changing one of the default INLA specifications. Inspecting the identifiable linear predictor MCMC and INLA coincide perfectly.

3.2 Analysis of the impact of family integration

Following the Durkheim tradition, one may ask whether explanatory variables on social integration can explain changes in suicide rates. The usual approach to study this is to look at the age-standardised male and female suicide rates. The matrix of pairwise correlations is shown

Table 3: Pair-wise Spearman correlations.

	Y1	Y2	X1	X2	X3
Male suicide rate (Y1)					
Female suicide rate (Y2)	0.86				
F-index (X1)	0.40	0.29			
Total male marriage rate (X2)	0.39	0.26	0.94		
Total female marriage rate (X3)	0.27	0.14	0.92	0.98	
Unemployment (X4)	-0.25	-0.27	-0.81	-0.75	-0.70

in Table 3. Note, that neither the male suicide rates nor the female suicide rates in Switzerland are negatively correlated to the F-index. The total gender-specific marriage rates are positively correlated to both rates. Also the correlation between the unemployment rate and suicide mortality has the contrary sign of what we expected and indicates an inverse effect. Note that some of the explanatory variables are highly correlated, for example unemployment and the F-index have a correlation of -0.81 ($p < 0.05$). However, since correlations alone might not be meaningful we continued fitting a linear model using ordinary least squares. We used separate models for male and female mortality rates including three regressors, namely an intercept, the rate of unemployment and one covariate on family integration. To detect autocorrelation of first order between two successive residuals we used the Durbin-Watson test (Harvey, 1990). For each analysis the Durbin-Watson statistic was smaller than the lower bound of 1.50 for 58 observations and three regressors, indicating positive autocorrelation. Hence, we included a one-year lagged dependent variable as a regressor to eliminate autocorrelation. In autoregressive models the Durbin-Watson test statistics tends to underestimate autocorrelation, so that we used Durbin's h statistic (Harvey, 1990). Durbin's h is a normally distributed variable, so that the h-statistics should be within -1.96 and 1.96, which was the case for all models.

Table 4 shows the regression estimates for all models. The first two columns show the results for the models including the F-index as marker for family integration. In contrast the second two columns show the results when including total gender-specific marriage rates. The lagged suicide rate is positively related to the dependent variable for both gender in all regressions.

Table 4: The effect of family integration on suicide for men and women in Switzerland from 1950 to 2007. Standard errors are given in parentheses.

	Suicide			
	Males	Females	Males	Females
Suicide rate lag	$0.880_{(0.061)}$	$0.769_{(0.078)}$	$0.875_{(0.061)}$	$0.719_{(0.079)}$
F-index	$-2.280_{(2.744)}$	$-2.518_{(1.502)}$	-	-
Tot. marriage rate (males)	-	-	$-0.037_{(0.025)}$	-
Tot. marriage rate (females)	-	-	-	$-0.041_{(0.015)}$
Unemployment	$-0.665_{(0.351)}$	$-0.461_{(0.199)}$	$-0.728_{(0.301)}$	$-0.505_{(0.164)}$
Constant	$6.065_{(3.080)}$	$4.972_{(1.787)}$	$7.659_{(3.192)}$	$7.231_{(2.048)}$
Adjusted R-squared	0.858	0.771	0.862	0.786
Durbin's h statistic	-1.49	-1.74	-1.73	-1.94

Inspecting the results of the models including the F-index we see that neither the male nor the female suicide rate is related to the F-index. Unemployment is significantly negatively related to the female suicide rate.

Turning to the regressions that include the total gender-specific marriage rate the unemployment rate is again significantly negatively related to male and female suicide rates. For males the coefficient of the total male marriage rate is -1.48 ($-0.037/0.025$) times its standard error. For females the coefficient of the corresponding female rate is -2.73 ($-0.041/0.015$) times its standard error. Hence for higher values of the total marriage rate the female suicide mortality decreases.

To be able to keep the age-specific structure of the data, so that all information can be used, we integrate the explanatory variables into the APC model and estimate the models using INLA. In the following we assume a time-constant effect of the rate of unemployment. Thus, we replace the period parameters with one linear effect of the unemployment rate and either a linear, quadratic or non-parametric effect of the F-index or the total gender-specific marriage rates. Modelling the covariate in a non-parametric fashion we assume a random walk of second order as non-parametric prior for the covariate effect. This is a natural prior as it models deviations from

Table 5: Log score and DIC of univariate APC models including either no covariate, or a time-constant effect of unemployment rate and either a time-constant (linear), quadratic or non-parametric effect of an explanatory variable on family integration.

	Females		Males	
	Log score	DIC	Log score	DIC
No covariate	**3.090**	**849.23**	**3.677**	**1001.25**
Time-constant linear effect				
F-index	3.158	930.03	3.849	1135.35
Tot. marriage rate	3.151	922.39	3.814	1113.61
Time-constant quadratic effect				
F-index	3.116	881.53	3.721	1042.10
Tot. marriage rate	3.152	923.71	3.811	1113.00
Random walk of second order on covariate				
F-index	3.107	870.07	3.697	1021.06
Tot. marriage rate	3.119	885.49	3.730	1049.77

a linear trend but reduces to the linear model if its variance goes to zero (Natario and Knorr-Held, 2003). In Natario and Knorr-Held (2003) an inverse gamma distribution with shape equal to 1 and scale equal to 0.00005 is proposed as prior for non-equally spaced covariates with an average distance equal to one. We used this prior and scaled each covariate on family integration appropriately. In addition, a sum-to-zero constraint is applied on the covariate effects to ensure identifiability of the intercept.

Table 5 shows the model choice criteria for all models. For both sexes the model without covariates is clearly classified as the best model, which indicates that the covariates we proposed cannot fully replace the period effects. However, note that assuming a non-parametric effect of the F-index is much better than the corresponding parametric formulations assuming a linear or quadratic effect. In addition, this model is also not so far away from the standard APC model for both sexes. Table 6 shows the parameter estimates of all models with parametric covariate effects. The unemployment rate has a slight negative effect on the log female suicide rate in all models and on the log male suicide rate in models with a linear effect of F-index or total marriage

Table 6: Parameter estimates (posterior median, 2.5% and 97.5% quantile) for models with parametric covariate effects.

	Females			Males		
	2.5% qu.	Median	97.5% qu.	2.5% qu.	Median	97.5% qu.
Time-constant F-index						
(Intercept)	-8.05	-7.85	-7.66	-7.27	-7.11	-6.95
F-index	-1.96	-1.65	-1.34	-1.42	-1.17	-0.92
Unemployment	-0.09	-0.07	-0.06	-0.07	-0.05	-0.04
Time-constant total marriage rate						
(Intercept)	-8.03	-7.85	-7.66	-7.10	-6.95	-6.80
Tot. marriage rate	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01
Unemployment	-0.06	-0.04	-0.02	-0.04	-0.03	-0.02
Time-constant quadratic F-index						
(Intercept)	-9.52	-9.15	-8.78	-9.11	-8.84	-8.57
F-index	1.58	2.69	3.79	3.82	4.65	5.47
F-index^2	-4.31	-3.45	-2.61	-5.27	-4.63	-4.00
Unemployment	-0.06	-0.04	-0.02	-0.02	-0.01	0.01
Time-constant quadratic total marriage rate						
(Intercept)	-9.23	-8.18	-7.13	-8.46	-7.78	-7.10
Tot. marriage rate	-0.03	-0.00	0.02	-0.01	0.01	0.03
Tot. marriage rate^2	-0.00	-0.00	0.00	-0.00	-0.00	-0.00
Unemployment	-0.06	-0.04	-0.02	-0.03	-0.02	-0.00

rate. The F-index has a negative linear effect for both males and females. Thus, increasing the F-index by 0.1 units reduces the female suicide risk by 15% ($= 100 \cdot (1 - \exp(-0.165))$). Modelling the F-index in a quadratic way $\beta_0 \cdot$F-index $+\beta_1 \cdot$F-index^2, the estimate of β_0 is positive while the estimate of β_1 is negative for both males and females. Figure 4 shows the parametric and non-parametric estimates for the F-index and the total male and female marriage rate. Assuming a quadratic and non-parametric effect for the F-index results in very similar estimates. In contrast for the total marriage rate the quadratic and linear effect are very similar. Regarding the model choice criteria presented in Table 5 the inclusion of the F-index is preferred compared to the inclusion of the total marriage rate.

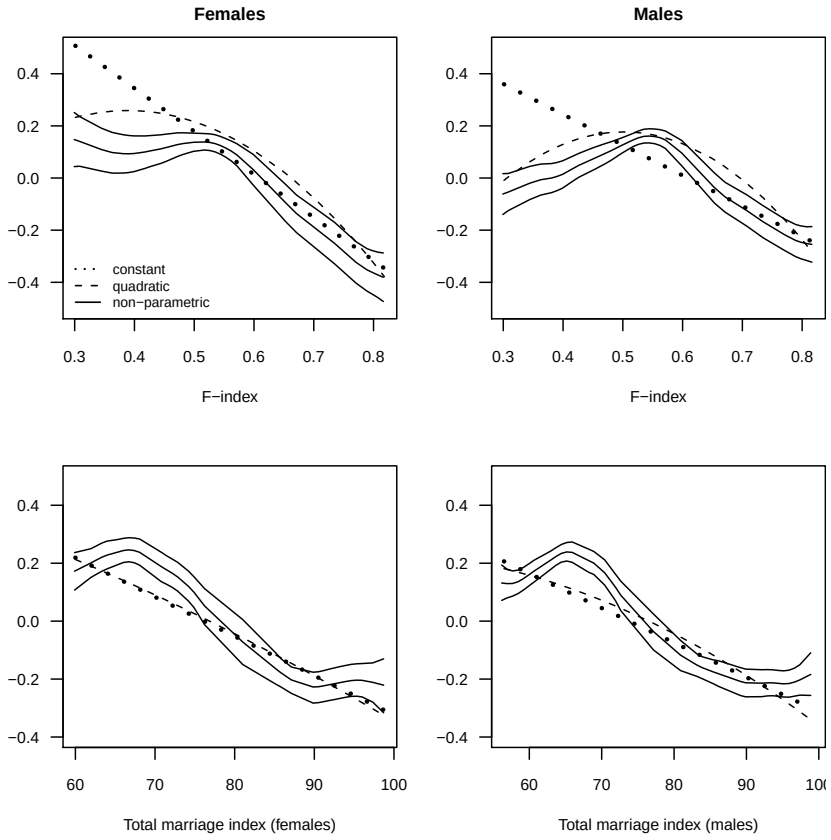

Figure 4: Estimated parametric (linear and quadratic) and non-parametric effects of covariates related to family integration. For the non-parametric trend 95% pointwise credible bands are shown.

Figure 4 shows that the effect of the F-index is very similar for males and females which suggests a joint analysis with correlated non-parametric F-index effects. We use the best-classified correlated multivariate APC model of Section 3.1, namely the "joint period effects" model with correlated age and cohort effects and correlated overdispersion parameters, and replace the period effects by a non-parametric effect of the F-index which is assumed to be correlated between males and females. Figure 5 shows the resulting age effects, cohort effects and covariate effects. The estimated time variables exhibit a similar pattern for males and females. Table 7 shows that the correlation estimates are, except for the cohort effects, very high and significantly different from zero. Thus, similar risk factors may act on age and overdispersion. In addition,

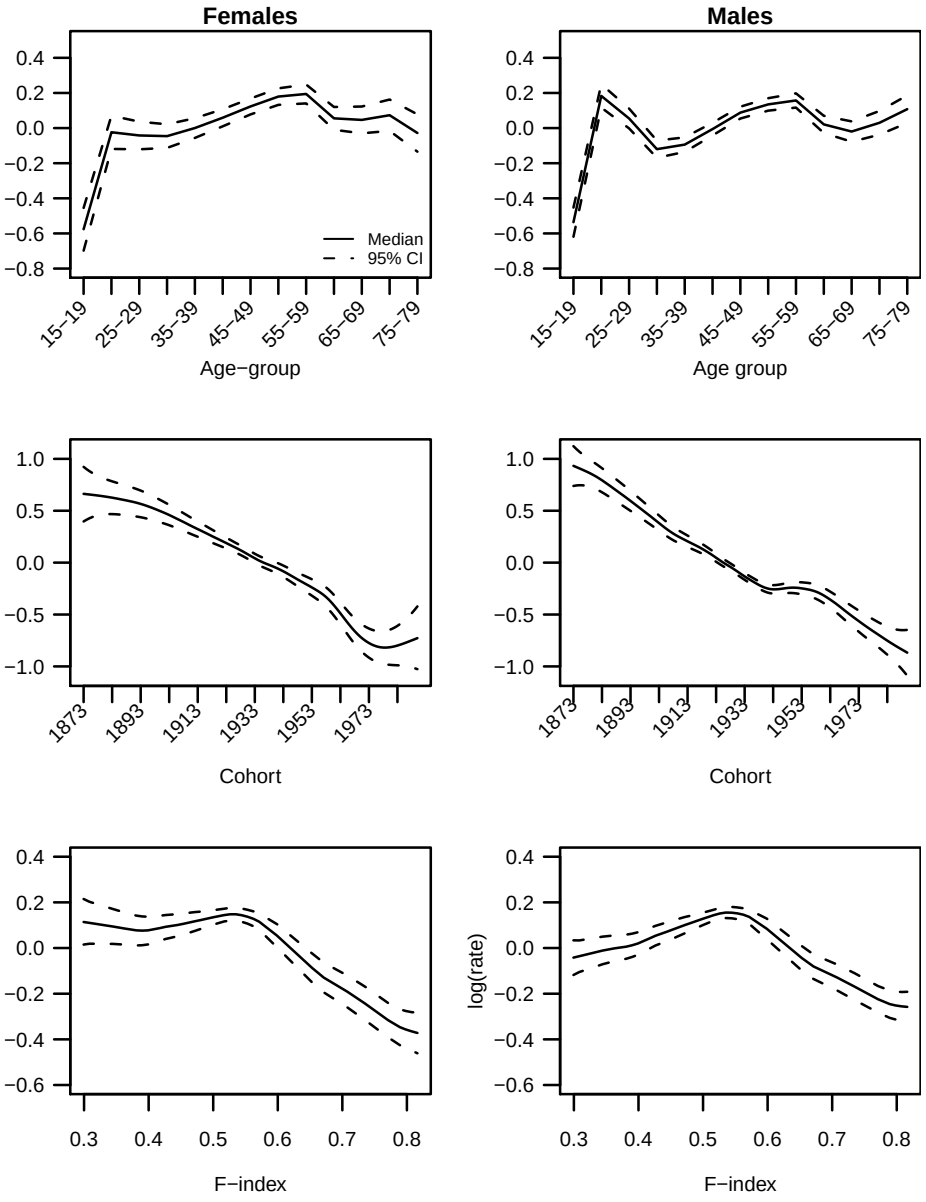

Figure 5: Estimated age effects, cohort effects and non-parametric effects of the F-index obtained from a multivariate AC model assuming correlation between each pair of gender-specific effects: age effects, covariate effects, cohort effects and overdispersion.

the F-index has a similar effect on males and females indicated by the high estimated correlation. Figure 5 shows that the age effects strongly increase from the youngest ages to the 20–24 years old persons. Then, for females, the effects slightly increase until the age of around 60 when

Table 7: Model choice criteria and correlation estimates obtained from INLA for the AC model with correlated **A**ge and **C**ohort effects, correlated **O**verdispersion and a correlated non-parametric effect of the F-index (third column). For comparison the corresponding model without correlation (second column) and the correlated model without covariate but joint period effects (first column) is given. The triple notation ${}_LP_U$ for the correlation parameters denotes the posterior median P with 2.5 per cent quantile L and 97.5 per cent quantile U.

		Period effects replaced by F-index	
	A,C,O	no correlation	A,F-index,C,O
Model choice			
Log Score	**3.386**	3.402	3.393
DIC	**1854.60**	1890.05	1870.26
Correlation coefficients			
Correlation age	${}_{0.50}0.82_{0.95}$	-	${}_{0.50}0.82_{0.95}$
Correlation F-index	-	-	${}_{0.50}0.94_{1.00}$
Correlation cohort	${}_{-0.61}+0.20_{+0.79}$	-	${}_{-0.61}+0.23_{+0.81}$
Correlation overdispersion	${}_{0.18}0.56_{0.84}$	-	${}_{0.26}0.63_{0.85}$

they start to decrease again. For males the age effects clearly drop from the 20–24 age group to the 25–29 years old. Similar to the females the effects slightly increase until the age of 60, then start decreasing. For the oldest ones the age effects increase again. The cohort effects show a negative slope for both sexes from the oldest to the youngest cohorts. In contrast to males, the cohort effects in women stop decreasing for those born at the early 1970s. Finally, the effect of the F-index on females stays almost constant for values between 0.3 and 0.55. For larger values, representing a higher degree of social integration, the effect decreases. The effect on males increases slightly from a value of 0.3 to a maximum of about 0.55 and decreases for higher values as well. The joint time-constant effect of the unemployment rate is estimated as -0.00 (95% CI: $-0.02, 0.01$). Thus the effect of the unemployment rate is not significantly different from zero.

Figure 6 shows the estimated relative risks of suicide of males compared to females. The estimates are very similar to those obtained in Figure 3, but slightly lower and with larger credible bands.

In Table 7 the log score and DIC estimate compared to the corresponding model without assuming correlation for any of the effects and compared to the model regarded best in Section 3.1 are given. The correlated version is clearly preferred. However, the standard correlated multivariate "joint period effects" model is classified as the best model.

4 Discussion

The results of the present multivariate age-period-cohort analysis confirm previous findings of strong gender-specific differences in suicide rates. Nevertheless the correlation estimates between gender-specific effects were, except for the cohort effects, very high. So it seems that similar risk factors act especially on age and overdispersion, which results in similar effect curves but on different overall levels. For all years and all age groups men have, compared to women, a three-fold risk to commit suicide. Elderly men and those between 16 and 25 show a especially high relative risk compared to their female peers in the same age group. This is quite surprising considering the higher number of suicide attempts of females compared to males. An explanation might be that males use more lethal methods like hanging or firearms, while the most frequent method of females is poisoning (Ajdacic-Gross *et al.*, 2008).

We further found a pronounced elevation for males born around 1970. Ajdacic-Gross *et al.* (2006) found in univariate APC models an inflexion point in cohort effects around 1970 for males as well, but not for females. Explaining this observed pattern is difficult. There should be risk factors related to cohorts born around 1970, but not to those born before or after 1970. Since the effect seems to be present especially in males and not females, also reflected in an insignificant correlation estimate, we suppose that the causal factor must occur in later life. This is plausible in the context of suicide. Relevant risk factors for suicide are complex, so a better understanding is necessary to explain the observed pattern. Stockard and O'Brien (2002) suggested that cohorts experiencing less social integration and having a large relative cohort size have higher suicide rates. However, in Switzerland the persons around 1964 are the largest cohort and therefore do not explain the peak for cohorts born after 1970.

We further explored whether family integration could explain gender-specific differences in suicide rates. We started with standard time-series analysis based on age-standardised rates. A significant influence was only found for the total female marriage rate on female suicide rate, indicating that with a higher total female marriage rate suicide rates of women decrease. Exploiting the age-specific structure of the data we applied univariate age-period cohort models and replaced the period effects by parametric and non-parametric effects of variables related to family integration. To adjust for confounding induced by the rate of unemployment we additionally included a linear effect of the unemployment rate. The inclusion of the so-called F-index measured as (marriage rate - divorce rate)/ (marriage rate + divorce rate) was preferred compared to the gender-specific total marriage rate. Higher values, corresponding to better integration, have a decreasing effect on suicide risk. Of note, the effect of the F-index on males and females is very similar. Thus, we used a multivariate APC model and replaced the period effects by a correlated non-parametric effect of the F-index. The estimated correlation was very high. A difference we found for women and men was, that median values of the F-index, interpretable as normally integrated, have a higher effect on male suicide than being worse integrated. As the computation of the F-index is only based on marriage and divorce rates, this marker might be not suitable to measure bad integration. Many males might be not married, but nevertheless well integrated and happy.

By means of model choice criteria the standard APC model without covariates was preferred. This indicates that the measures of social integration we used cannot fully replace the period effects. If we had analysed only data from 1950 to 1980, we probably would have found a strong linear dependence between the F-index and the suicide rates, because from 1950 to 1970 the F-index value was almost constantly at 0.8. This indicates a high degree of social integration. In the same time-period suicide rates were strongly decreasing for both sexes. From 1970 to 1980 the F-index dropped and almost inversely the suicide rates increased. However after 1980 this linear relationship has vanished.

A problem of the present analysis is that it is very difficult to decide how to measure social or

family integration. There are several proposals and all of them are controversially discussed. In addition, it is not completely clear how to properly include information on explanatory variables in the APC model (Knorr-Held and Rainer, 2001). Since social integration might not be the main risk factor for committing suicide, it could be necessary to adjust for further confounding variables, for example mental disease or religiousness. Note that it is also possible to include more covariates in a non-parametric fashion. In our analysis the two markers on social integration are strongly correlated, so that in this cases it is not meaningful to include both. Alternatively, the period effects could be kept in the model, but then the identifiability problem well known for APC models remains.

It could also be that the covariates on social integration exert a lagged influence on suicide mortality. For example, Wasserman (1984) determined a lag of nine months of the influence of divorce on suicide reported in the United States. Including a lagged covariate into a Bayesian APC model is particularly attractive, because then projections of suicide rates can be generated for future periods without any parametric assumptions. More research would be necessary to explore whether such a lagged effect is also present for Swiss suicide data. However, it is difficult to determine an exact time-lag because the separation of couples starts much earlier than the divorce proceeding is completed. Hence, the exact time-point which is relevant for suicide behaviour is difficult to determine.

Acknowledgements

This project was supported by the Swiss National Science Foundation.

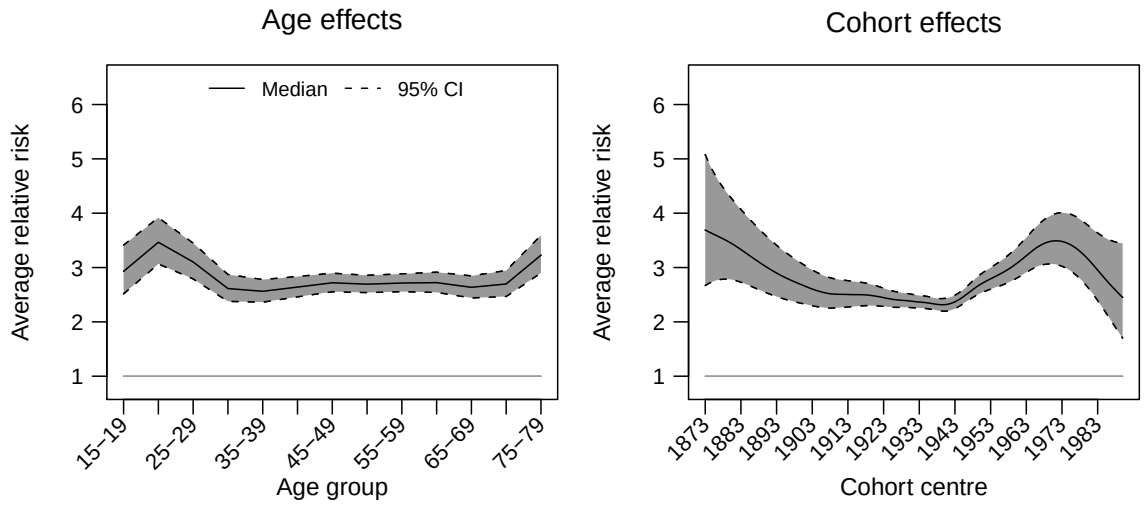

Figure 6: Estimated relative risk of suicide for males compared to females obtained from a multivariate model assuming correlation between each pair of gender-specific effects: age effects, covariate effects, cohort effects and overdispersion.

References

Ahmad, O., Boschi-Pinto, C., Lopez, A. D., Murray, C. J. L., Lozano, R. and Inoue, M. (2001). Age standardization of rates: a new WHO standard, *GPE Discussion Paper Series*, Geneva. (GPE Discussion Paper No. 31).

Ajdacic-Gross, V., Bopp, M., Gostynski, M., Lauber, C., Gutzwiller, F. and Rössler, W. (2006). Age-period-cohort analysis of Swiss suicide data, 1881-2000, *European Archives of Psychiatry and Clinical Neuroscience* **256**(4): 207–214.

Ajdacic-Gross, V., Weiss, M. G., Ring, M., Hepp, U., Bopp, M., Gutzwiller, F. and Rössler, W. (2008). Methods of suicide: International suicide patterns derived from the WHO mortality database, *Bulletin of the World Health Organization* **86**: 726 – 732.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Baysesian computation of stochastic-systems, *Statistical Science* **10**(1): 3–41.

Breault, K. and Barkey, K. (1982). A comparative analysis of Durkheim's theory on egoistic suicide, *Sociological Quarterly* **23**(3): 321–331.

Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics, *Biometrics* **42**(4): 693–712.

Brown, C. C. and Kessler, L. G. (1988). Projections of lung cancer mortality in the United States: 1985-2025, *Journal of the National Cancer Institute* **80**(1): 43–51.

Calot, G. (1998). *Two centuries of Swiss demographic history. Graphic album of the 1860-2050 period*, Swiss Federal Statistical Office, Neuchâtel.

Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models, *Statistics in Medicine* **6**(4): 469–481.

Durkheim, E. (1897). *Suicide: A Study in Sociology*, Free Press, New York. 1966.

Etzersdorfer, E., Piribauer, F. and Sonneck, G. (1996). Sex differential for suicide among Austrian age cohorts, *Acta Psychiatrica Scandinavica* **93**(4): 240–245.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* **102**(477): 359–378.

Granizo, J. J., Guallar, E. and Rodriguez-Artalejo, F. (1996). Age-period-cohort analysis of suicide mortality rates in Spain, 1959-1991, *International Journal of Epidemiology* **25**(4): 814–820.

Gunnell, D., Middleton, N., Whitley, E., Dorling, D. and Frankel, S. (2003). Influence of cohort effects on patterns of suicide in England and Wales, 1950-1999, *British Journal of Psychiatry* **182**: 164–170.

Harvey, A. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*, reprinted edn, Cambridge University Press, Cambridge.

Heuer, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines, *Biometrics* **53**(1): 161–177.

Knorr-Held, L. and Rainer, E. (2001). Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction, *Biostatistics* **2**(1): 109–129.

Lester, D. (1986). The interaction of divorce, suicide and homicide, *Journal of Divorce* **9**(3): 103–109.

Lester, D. (1990). The effect of the detoxification of domestic gas in Switzerland on the suicide rate, *Acta Psychiatrica Scandinavica* **82**(5): 383–384.

Lester, D. (1994). Domestic integration and suicide in 21 nations, 1950-1985, *International Journal of Comparative Sociology* **35**: 131–137.

Levi, F., La Vecchia, C., Lucchini, F., Negri, E., Saxena, S., Maulik, P. K. and Saraceno, B. (2003). Trends in mortality from suicide, 1965-99, *Acta Psychiatrica Scandinavica* **108**(5): 341–349.

Natario, I. and Knorr-Held, L. (2003). Non-parametric ecological regression and spatial variation, *Biometrical Journal* **45**(6): 670–688.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison, *Biostatistics* **9**(3): 523–539.

Riebler, A. (2010). *Multivariage age-period-cohort models*, PhD thesis, University of Zurich.

Riebler, A. and Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models, *Biostatistics* **11**(1): 57–69.

Riebler, A., Held, L. and Rue, H. (2010). Correlated multivariate age-period-cohort models, *Technical report*, University of Zurich.

Rossow, I. (1993). Suicide, alcohol, and divorce - aspects of gender and family integration, *Addiction* **88**(12): 1659–1665.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society - Series B* **71**: 319–392.

Shah, A. and De, T. (1998). Suicide and the elderly, *International Journal of Psychiatry in Clinical Practice* **2**: 3–17.

Snowdon, J. and Hunt, G. E. (2002). Age, period and cohort effects on suicide rates in Australia, 1919-1999, *Acta Psychiatrica Scandinavica* **105**(4): 265–270.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R. and van der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society - Series B* **64**: 583–616.

Stack, S. (1987). The effect of divorce on suicide - Some methodological problems, *Journal of Marriage and Family* **49**(1): 205–206.

Stack, S. (1989). The impact of divorce on suicide in Norway, 1951-1980, *Journal of Marriage and Family* **51**(1): 229–238.

Stack, S. (1990a). Divorce, suicide, and the mass-media - An analysis of differential identification, 1948-1980, *Journal of Marriage and Family* **52**(2): 553–560.

Stack, S. (1990b). New micro-level data on the impact of divorce on suicide, 1959-1980 - A test of 2 theories, *Journal of Marriage and Family* **52**(1): 119–127.

Stack, S. (1992). The effect of divorce on suicide in Finland - A time-series analysis, *Journal of Marriage and Family* **54**(3): 636–642.

Stockard, J. and O'Brien, R. M. (2002). Cohort effects on suicide rates: International variations, *American Sociological Review* **67**(6): 854–872.

Surault, P. (1992). Marriage rate, divorce rate and suicide rate - Closing schisms, *Population* **47**(4): 1042–1044.

The Swiss Federal Statistical Office (2009). `http://www.bfs.admin.ch/bfs/portal/de/index/infothek/lexikon/bienvenue___login/blank/zugang_lexikon.Document.67153.xls`. accessed on 14.5.2010.

Wasserman, I. M. (1984). A longitudinal analysis of the linkage between suicide, unemployment, and marital dissolution, *Journal of Marriage and Family* **46**(4): 853–859.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* **57**(298): 348–368.

APPENDIX I

Correlated GMRF priors for multivariate age-period-cohort models

Andrea Riebler, Leonhard Held & Håvard Rue

Paper published in *Proceedings of the 25th International Workshop on Statistical Modelling. Glasgow, July 5-9, 2010, Adrian W. Bowman (editor), pp. 455–460.*

Correlated GMRF priors for multivariate age-period-cohort models

Andrea Riebler[1], Leonhard Held[1] and Håvard Rue[2]

[1] Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland;
Email: {andrea.riebler, leonhard.held}@ifspm.uzh.ch,
[2] Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway; Email: havard.rue@math.ntnu.no

Abstract: Multivariate age-period-cohort models have recently been proposed for the analysis of heterogeneous time trends. For a fully Bayesian analysis, Gaussian Markov random field (GMRF) priors are typically used. However, standard GMRF priors do not account for a potential dependence between outcomes. We present an extended approach based on correlated smoothing priors and correlated overdispersion parameters. Algorithmic routines are based on either Markov chain Monte Carlo or integrated nested Laplace approximations. Results are discussed for data on female mortality in Denmark and Norway and compared by means of DIC, proper scoring rules and the marginal likelihood.

Keywords: Bayesian analysis; Gaussian Markov random field; INLA; Multivariate age-period-cohort model; Uniform correlation matrix.

1 Introduction

Age-period-cohort (APC) models are used to analyse mortality or disease counts stratified by age and period. For the case in which rates are available for multiple health outcomes multivariate APC models have been proposed, see e. g. Jacobsen et al. (2004) or Riebler and Held (2010). A joint analysis may borrow strength from a set of shared effects, for example, the age effects while possibly identifying different period or cohort effects. Within a Bayesian setting, typically, both overdispersion parameters and smoothing priors on the time trends are assumed to be independent across outcomes. Hence, a potential dependence between the outcomes is not captured.
We present an extended approach based on correlated overdispersion parameters and correlated smoothing priors. The latter involves a Kronecker product structure composed of the inverse of a uniform correlation matrix and the precision matrix of the univariate second-order random walk (RW2). Fully Bayesian inference is conducted by either Markov chain Monte Carlo (MCMC) or integrated nested Laplace approximations (INLA) (Rue et al., 2009). The methodology will be applied to mortality rates among

Danish and Norwegian women and models will be compared based on proper scoring rules (Gneiting and Raftery, 2007), the well-known deviance information criterion (DIC) and the marginal likelihood.

2 The correlated multivariate APC model

Let n_{ijs} denote the number of persons under risk in age group i ($i = 1, \ldots, I$), period j ($j = 1, \ldots, J$) and health outcome s ($s = 1, \ldots, S$). We assume that the number of disease cases or deaths y_{ijs} follows a Poisson distribution with mean $n_{ijs}\lambda_{ijs}$, where in the most general formulation

$$\eta_{ijs} = \log(\lambda_{ijs}) = \mu_s + \theta_{is} + \phi_{js} + \psi_{ks}. \quad (1)$$

Here, μ_s is the outcome-specific intercept, and θ_{is}, ϕ_{js} and ψ_{ks} are outcome-specific age, period and cohort effects, respectively. The cohort index k depends on age index i and period index j and is defined as $M \times (I - i) + j$ where M is the ratio of the widths of the age group and period intervals. Simpler models can be obtained, for example by assuming shared period effects. Then, the linear predictor is

$$\eta_{ijs} = \log(\lambda_{ijs}) = \mu_s + \theta_{is} + \phi_j + \psi_{ks}. \quad (2)$$

Since we are in a Bayesian context all parameters are treated as random variables and prior distributions need to be assigned. We use a flat prior for each μ_s and assume that second differences of shared time effects, here the period effects, are independent Gaussian variables. For outcome-specific time effects, here the age and cohort effects, we use a correlated GMRF prior with precision matrix $\mathbf{P} = \mathbf{C}^{-1} \otimes \mathbf{R}$. Here, $\mathbf{C}^{-1}$ is the inverse of the $S \times S$ uniform correlation matrix $\mathbf{C} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$, where ρ denotes the correlation parameter, $\mathbf{I}$ the identity matrix and $\mathbf{J}$ is a matrix of ones, and $\mathbf{R}$ is the precision matrix of the univariate RW2 (see Rue and Held, 2005, page 110). This formulation corresponds to a multivariate RW2 with correlated increments. Note that we assign to each time-scale an individual precision and in the case of outcome-specific effects an individual correlation parameter. Sum-to-zero constraints are assumed for each parameter vector, in (2) θ_s, ϕ and ψ_s with $s = 1, \ldots, S$.

To adjust for unobserved heterogeneity we introduce further outcome-specific variables z_{ijs} into the linear predictor (1). Typically, these overdispersion parameters are assumed to be independent Gaussian variables with mean zero and unknown variance (Besag et al., 1995). We propose correlated overdispersion parameters and set $z_{ij} = (z_{ij1}, \ldots, z_{ijS})^\top \sim \mathrm{N}(0, \tau_z^{-1}\mathbf{C})$ for all i and j, where τ_z denotes the precision of the overdispersion.

All of the up to eight hyperparameters (four precisions and up to four correlations) are treated as unknown. Suitable gamma-hyperpriors are assigned to the precisions. To each correlation ρ we apply Fisher's z-transformation

$$\tilde{\rho} = \log\left(\frac{1 + \rho}{1 - \rho}\right), \quad -1 < \rho < 1,$$

and assign a Gaussian prior with mean zero and variance 0.2^{-1} to $\tilde{\rho}$, corresponding to a U-shaped prior for correlation ρ. To ensure positive definiteness of $\mathbf{C}$ the additional constraint $\rho > -1/(S-1)$ is required.

3 Implementation

Algorithmic routines based on MCMC were implemented in the low-level programming language `C` using the `GMRFLib` library (Rue and Held, 2005). Following Besag et al. (1995), we reparameterised the model from z_{ijs} to η_{ijs} to obtain multivariate normal full conditional distributions for the intercepts and time effects. Block updating allows the proper incorporation of the sum-to-zero constraints for the time effects. For the precisions also Gibbs sampling is used. The vector $\eta_{ij} = (\eta_{ij1}, \ldots, \eta_{ijS})^\top$ has a non-standard distribution. It is updated using multivariate Metropolis-Hastings steps with a GMRF proposal distribution based on a second-order Taylor approximation of the log-likelihood. For the correlation parameters Metropolis-Hastings updates based on a random walk proposal are used, such that acceptance rates around 40% are achieved.
An attractive and fast alternative to MCMC in the class of latent Gaussian random field models is INLA (Rue et al., 2009). This approach directly computes very accurate approximations to the posterior marginal distributions, so that MCMC sampling becomes redundant. We included a new option in the `inla` programme to correlate a wide range of latent GMRF models based on a uniform correlation structure. The methodology can be applied using the `R`-package `INLA` (see `www.r-inla.org`). Here, we use the `INLA` package built on 09.04.2010.

4 Model choice

The DIC is frequently used for model comparison. It is the sum of the posterior saturated deviance $\bar{D}$, a measure for model fit, and the effective number of parameters p_D, a measure for model complexity. Within both MCMC and INLA, estimates for DIC can be calculated. However, for hierarchical models with many random effects, as in (1) with included overdispersion parameters, the use of DIC has recently been criticised (Plummer, 2008).
An alternative are proper scoring rules, e.g. the mean Dawid-Sebastiani score (Riebler and Held, 2010). To account for the correlation potentially present in multiple outcomes and captured by using correlated GMRF priors, this score needs to be adapted. We denote this generalised form as multivariate mean David-Sebastiani score $\overline{\text{MDSS}}$. Within MCMC we used approximate leave-one-block-out cross-validation based on replicating the vector $\eta_{ij} = (\eta_{ij1}, \ldots, \eta_{ijS})^\top$ and subsequently the observation vector $\mathbf{y}_{ij} = (y_{ij1}, \ldots, y_{ijS})^\top$ (Marshall and Spiegelhalter, 2003). These replicated

data points can now be used to calculate the $\overline{\text{MDSS}}$ as:

$$\overline{\text{MDSS}} = \frac{1}{IJ}\sum_{i,j}\left[\left(\mathbf{y}_{ij} - \overline{\mathbf{y}_{ij}^{\text{rep}}}\right)^{\top}\{\boldsymbol{\Sigma}_{ij}^{\text{rep}}\}^{-1}\left(\mathbf{y}_{ij} - \overline{\mathbf{y}_{ij}^{\text{rep}}}\right) + \log\left|\boldsymbol{\Sigma}_{ij}^{\text{rep}}\right|\right]$$

where $\overline{\mathbf{y}_{ij}^{\text{rep}}} = (\overline{y_{ij1}^{\text{rep}}}, \ldots, \overline{y_{ijS}^{\text{rep}}})^{\top}$ and $\overline{y_{ijs}^{\text{rep}}}$ is the mean of the N replicated observation samples $\mathbf{y}_{ijs}^{\text{rep}} = (y_{ijs_{(1)}}^{\text{rep}}, \ldots, y_{ijs_{(N)}}^{\text{rep}})^{\top}$. Analogously, $\boldsymbol{\Sigma}_{ij}^{\text{rep}}$ represents the empirical covariance matrix of $(\mathbf{y}_{ij1}^{\text{rep}}, \ldots, \mathbf{y}_{ijS}^{\text{rep}})^{\top}$.
Furthermore, INLA returns an estimate of the log marginal likelihood $\log(p(\mathbf{y}))$. Usually the marginal likelihood is difficult to use for hierarchical GMRF models in which the prior is improper (here because of the RW2). However, for comparing models that only differ by the inclusion of correlation but have the same underlying first-level structure, e.g. (2), $\log(p(\mathbf{y}))$ can be used for model choice.

5 Mortality of Danish and Norwegian women

We analyse data on overall mortality, aggregated to 5-year age group and period intervals (i.e. $M = 1$), for all Danish and Norwegian women aged 0-84 years during the period 1960-1999 (Jacobsen et al., 2004). In an uncorrelated multivariate APC analysis Riebler and Held (2010) classified the aPc_z model with separate age and cohort effects but joint period effects as best. Here, we compare the aPc_z model with independent RW2 priors for θ_s, ϕ and ψ_s, $s = 1, 2$, to three different correlated models. Either age and cohort effects ($\text{a}^{\star}\text{Pc}_z^{\star}$ model), or the overdispersion parameters ($\text{aPc}_{z^{\star}}$ model) are correlated. Both correlated time and overdispersion parameters are specified in model $\text{a}^{\star}\text{Pc}_{z^{\star}}^{\star}$. For all models MCMC and INLA produce virtually identical results. The posterior correlation estimates of the $\text{a}^{\star}\text{Pc}_{z^{\star}}^{\star}$ model clearly indicate the dependence present between the outcomes, compare Figure 1. Table 1 shows the model choice criteria obtained by MCMC and INLA for all models. The $\text{a}^{\star}\text{Pc}_{z^{\star}}^{\star}$ model is clearly preferred.
Figure 2 shows estimates of relative risks for the $\text{a}^{\star}\text{Pc}_{z^{\star}}^{\star}$ model together with estimates of the uncorrelated aPc_z model. The estimates of the correlated model are smoother and the credible regions are smaller. For some interpretation of the relative risks see Riebler and Held (2010).

6 Conclusion

We proposed the use of correlated GMRF priors for multivariate age-period-cohort models and implemented these models based on a uniform correlation structure in MCMC and INLA. We illustrated the methodology on female mortality in Norway and Denmark and received virtually identical results with both approaches, MCMC and INLA. A correlated

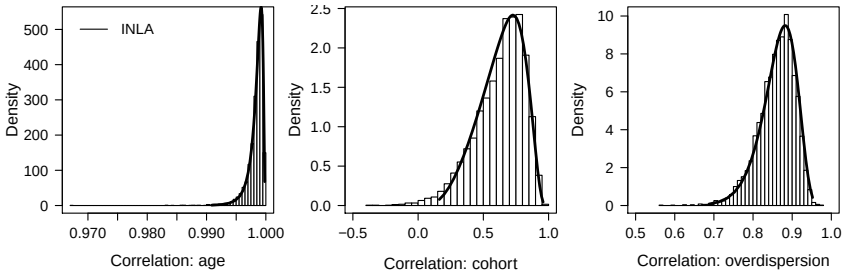

FIGURE 1. Posterior correlation estimates of the a⋆Pc⋆$_{z^\star}$ model. Approximated posterior marginals of INLA and corresponding histograms based on 5000 MCMC samples after 20 000 burn-in iterations and a thinning of 20 are shown.

TABLE 1. Model choice criteria obtained from MCMC and INLA. For both approaches deviance summaries are given. In addition, the multivariate mean Dawid-Sebastiani score $\overline{\text{MDSS}}$ and the log marginal likelihood are shown. (For each measure the best value is indicated in bold.)

	aPc$_z$	a*Pc$^\star_z$	aPc$_{z^\star}$	a*Pc$^\star_{z^\star}$
MCMC model choice				
$\overline{\text{D}}$	301.2	304.4	293.5	**292.4**
p_D	201.1	194.4	183.6	**176.4**
DIC	502.2	498.8	477.1	**468.8**
$\overline{\text{MDSS}}$	19.79	19.71	19.40	**19.39**
INLA model choice				
$\overline{\text{D}}$	301.6	304.6	293.6	**292.5**
p_D	201.1	194.2	183.7	**176.3**
DIC	502.7	498.9	477.3	**468.8**
log($p(\mathbf{y})$)	-1799.6	-1765.6	-1776.3	**−1741.1**

model structure lead to more precise relative risk estimates and was clearly preferred in this application.
However, benefits of correlated multivariate APC models might not only be in terms of model choice criteria and the improved precision of relative risk estimates. When projecting e.g. mortality rates of one health outcome a correlated joint analysis with a set of comparable outcomes may borrow strength from these and thus lead to more accurate projections.

Acknowledgments: This work received support from the Swiss National Science Foundation and the Research Council of Norway.

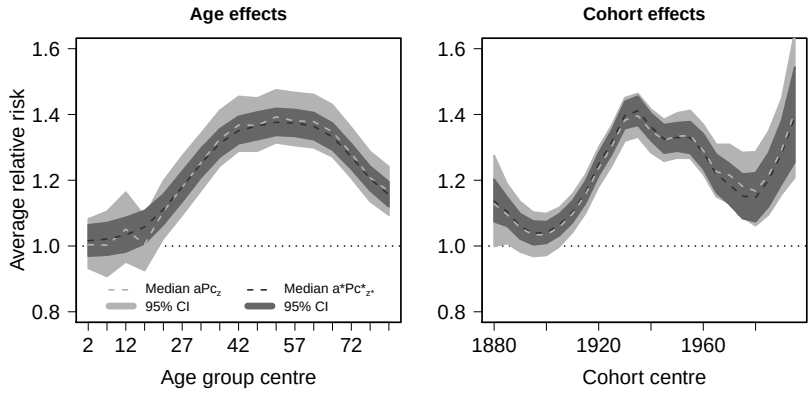

FIGURE 2. Average relative risk of death for Danish compared with Norwegian women analysed by the aPc_z and $\text{a}^\star\text{Pc}^\star_{z\star}$ model.

References

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, **10**, 3-41.

Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359-378.

Jacobsen, R., von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - What makes Denmark different? *European Journal of Epidemiology*, **19**, 117-121.

Marshall, E.C. and Spiegelhalter, D.J. (2003). Approximate cross-validatory predictive checks in disease-mapping methods. *Statistics in Medicine*, **22**, 1649-1660.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, **9**, 523-539.

Riebler, A., and Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models. *Biostatistics*, **11**, 57-69.

Rue, H., and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman & Hall/CRC Press.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319-392.

APPENDIX II

Full conditional distributions

First, we will derive the full conditional distributions for the ordinary multivariate APC model. Then, we will generalise them to the correlated case. Note that we continue using the notation of the introduction. By adding additional overdispersion parameters accounting for unobserved heterogeneity, the implementation of APC models is considerably simplified, as the model formulation can be reparameterised according to Besag *et al.* (1995). The advantage of the reparameterised model is that the full conditional distributions for all latent parameters are now multivariate Gaussian, so updating can be performed with a simple (multivariate) Gibbs step (Knorr-Held and Rue, 2002). We use a block update for each set of age, period and cohort effects and single site updates for all hyperparameters. The full conditional distribution of the linear predictor as well the update of the linear predictor is discussed in the Introduction, Section 6.2.

1 Ordinary multivariate age-period-cohort models

We first present the full conditional distributions for the components of the latent field. Subsequently, we present those of the precision and correlation parameters.

1.1 Latent field

For simplicity, let the precision parameters be stratum-independent, i.e. there is only one precision for each time scale. The derivation of the full conditional distributions in the case of stratum-dependent precision parameters proceeds in an analogous way.

Using a prior based on second differences, the full conditional distribution for separate age effects $\boldsymbol{\theta}_r$, $r = 1, \ldots, R$, is given by:

$$
\begin{aligned}
f(\boldsymbol{\theta}_r|\cdot) &\propto f(\boldsymbol{\xi}|\cdot) f(\boldsymbol{\theta}_r|\kappa_\theta) \\
&\propto \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=3}^{I}(\theta_{ir} - 2\theta_{(i-1)r} + \theta_{(i-2)r})^2\right) \\
&\qquad \cdot \exp\left(-\frac{\kappa_z}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}(\xi_{ijr} - (\mu_r + \theta_{ir} + \varphi_{j(r)} + \psi_{k(r)}))^2\right) \\
&= \exp\left[-\frac{1}{2}\boldsymbol{\theta}_r^\top \mathbf{R}_\theta^{(2)} \boldsymbol{\theta}_r - \frac{\kappa_z}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}(\theta_{ir} - \underbrace{[\xi_{ijr} - \mu_r - \varphi_{j(r)} - \psi_{k(r)}]}_{m_{ijr}^\theta})^2\right] \\
&= \exp\left[-\frac{1}{2}\boldsymbol{\theta}_r^\top \mathbf{R}_\theta^{(2)} \boldsymbol{\theta}_r - \frac{\kappa_z}{2}\sum_{j=1}^{J}\begin{pmatrix} \theta_{1r} - m_{1jr}^\theta \\ \theta_{2r} - m_{2jr}^\theta \\ \vdots \\ \theta_{Ir} - m_{Ijr}^\theta \end{pmatrix}^\top \mathbf{I} \begin{pmatrix} \theta_{1r} - m_{1jr}^\theta \\ \theta_{2r} - m_{2jr}^\theta \\ \vdots \\ \theta_{Ir} - m_{Ijr}^\theta \end{pmatrix}\right]
\end{aligned}
$$

$\boldsymbol{\theta}_r|\cdot \sim \mathcal{N}(\mathbf{E}_\theta^{-1}\mathbf{e}_{\theta_r}, \mathbf{E}_\theta^{-1})$ where

$$
\mathbf{E}_\theta = \mathbf{R}_\theta^{(2)} + \kappa_z J \mathbf{I} \qquad\qquad \mathbf{e}_{\theta_r} = \kappa_z \cdot \left(\sum_{j=1}^{J} m_{1jr}^\theta, \ldots, \sum_{j=1}^{J} m_{Ijr}^\theta\right)^\top,
$$

with $m^{\theta}_{ijr} = \xi_{ijr} - \mu_r - \varphi_{j(r)} - \psi_{k(r)}$. Using a random walk of first order the precision matrix $\mathbf{R}^{(2)}_{\theta}$ is substituted by $\mathbf{R}^{(1)}_{\theta}$. If joint age effects $\boldsymbol{\theta}$ are assumed the full conditional distribution is defined as $\boldsymbol{\theta}|. \sim \mathcal{N}(\mathbf{E}^{-1}_{\theta}\mathbf{e}_{\theta}, \mathbf{E}^{-1}_{\theta})$ with

$$\mathbf{E}_{\theta} = \mathbf{R}^{(2)}_{\theta} + \kappa_z RJ\mathbf{I} \qquad \mathbf{e}_{\theta} = \kappa_z \left(\sum_{r=1}^{R}\sum_{j=1}^{J} m^{\theta}_{1jr}, \ldots, \sum_{r=1}^{R}\sum_{j=1}^{J} m^{\theta}_{Ijr} \right)^{\top}.$$

Analogously, the full conditional distribution for each $\boldsymbol{\varphi}_r$ using separate period effects is $\boldsymbol{\varphi}_r|. \sim \mathcal{N}(\mathbf{E}^{-1}_{\varphi}\mathbf{e}_{\varphi_r}, \mathbf{E}^{-1}_{\varphi})$ where

$$\mathbf{E}_{\varphi} = \mathbf{R}^{(2)}_{\varphi} + \kappa_z I\mathbf{I} \qquad \mathbf{e}_{\varphi_r} = \kappa_z \cdot \left(\sum_{i=1}^{I} m^{\varphi}_{i1r}, \ldots, \sum_{i=1}^{I} m^{\varphi}_{iJr} \right)^{\top},$$

with $m^{\varphi}_{ijr} = \xi_{ijr} - \mu_r - \theta_{i(r)} - \psi_{k(r)}$. For common period effects $\boldsymbol{\varphi}$ the full conditional distribution is $\boldsymbol{\varphi}|. \sim \mathcal{N}(\mathbf{E}^{-1}_{\varphi}\mathbf{e}_{\varphi}, \mathbf{E}^{-1}_{\varphi})$ where

$$\mathbf{E}_{\varphi} = \mathbf{R}^{(2)}_{\varphi} + \kappa_z RI\mathbf{I} \qquad \mathbf{e}_{\varphi} = \kappa_z \left(\sum_{r=1}^{R}\sum_{i=1}^{I} m^{\varphi}_{i1r}, \ldots, \sum_{r=1}^{R}\sum_{i=1}^{I} m^{\varphi}_{iJr} \right)^{\top}.$$

Using separate cohort effects $\boldsymbol{\psi}_r$ the full conditional distribution is $\boldsymbol{\psi}_r|. \sim \mathcal{N}(\mathbf{E}^{-1}_{\psi}\mathbf{e}_{\psi_r}, \mathbf{E}^{-1}_{\psi})$ with

$$\begin{aligned}
\mathbf{E}_{\psi} &= \mathbf{R}^{(2)}_{\psi} + \kappa_z \mathrm{diag}(q_1, \ldots, q_K) \\
q_k &= \sum_{i,j:k(i,j)=k} 1 \qquad \text{with } k = 1, \ldots, K \\
\mathbf{e}_{\psi_r} &= \kappa_z \cdot \left(\sum_{k(i,j)=1} m^{\psi}_{ijr}, \ldots, \sum_{k(i,j)=K} m^{\psi}_{ijr} \right)^{\top} \\
m^{\psi}_{ijr} &= \xi_{ijr} - \mu_r - \theta_{i(r)} - \varphi_{j(r)}.
\end{aligned}$$

While for common cohort effects $\boldsymbol{\psi}$ the full conditional distribution is $\boldsymbol{\psi}|. \sim \mathcal{N}(\mathbf{E}^{-1}_{\psi}\mathbf{e}_{\psi}, \mathbf{E}^{-1}_{\psi})$ with

$$\begin{aligned}
\mathbf{E}_{\psi} &= \mathbf{R}^{(2)}_{\psi} + \kappa_z \mathrm{diag}(q_1, \ldots, q_K) \\
q_k &= \sum_{r=1}^{R} \sum_{i,j:k(i,j)=k} 1 \qquad \text{with } k = 1, \ldots, K \\
\mathbf{e}_{\psi} &= \kappa_z \left(\sum_{r=1}^{R} \sum_{k(i,j)=1} m^{\psi}_{ijr}, \ldots, \sum_{r=1}^{R} \sum_{k(i,j)=K} m^{\psi}_{ijr} \right)^{\top},
\end{aligned}$$

with $m^{\psi}_{ijr} = \xi_{ijr} - \mu_r - \theta_{i(r)} - \varphi_{j(r)}$.

For each stratum $r = 1, \ldots, R$ a separate intercept μ_r with a flat prior $p(\mu_r) \propto$ const. is used, so that the full conditional distribution is a univariate normal distribution: $\mu_r|. \sim \mathcal{N}(E^{-1}_{\mu}e_{\mu_r}, E^{-1}_{\mu})$

where

$$
\begin{aligned}
E_\mu &= \kappa_z IJ \\
e_{\mu_r} &= \kappa_z \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ijr}^{\mu} \\
m_{ijr}^{\mu} &= \xi_{ijr} - \theta_{i(r)} - \varphi_{j(r)} - \psi_{k(r)}.
\end{aligned}
$$

1.2 Hyperparameters

In the case of ordinary multivariate APC models there are only precision parameters. As prior distributions for the precision parameters gamma distributions $\mathrm{Ga}(a, b)$ are used. We exemplify the derivation of the full conditionals for precision κ_θ. Similarly, the full conditional distributions for $\kappa_\varphi, \kappa_\psi$ and κ_z follow. Using a random walk of second order and assuming separate age effects, the full conditional for κ_θ is given by

$$
\begin{aligned}
f(\kappa_\theta|\cdot) &\propto \prod_{r=1}^{R} f(\boldsymbol{\theta}_r|\kappa_\theta) f(\kappa_\theta) \\
&\propto \prod_{r=1}^{R} \kappa_\theta^{\frac{I-2}{2}} \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=3}^{I}(\theta_{ir} - 2\theta_{(i-1)r} + \theta_{(i-2)r})^2\right) \cdot \kappa_\theta^{a_\theta - 1} \exp(-b_\theta \kappa_\theta) \\
&= \prod_{r=1}^{R} \kappa_\theta^{a_\theta + \frac{I-2}{2} - 1} \exp\left(-\kappa_\theta\left(b_\theta + \frac{1}{2}\sum_{i=3}^{I}(\theta_{ir} - 2\theta_{(i-1)r} + \theta_{(i-2)r})^2\right)\right),
\end{aligned}
$$

which is gamma distributed:

$$
\kappa_\theta|\cdot \sim \mathrm{Ga}\left(a_\theta + \frac{R(I-2)}{2}, b_\theta + \frac{1}{2}\sum_{r=1}^{R}\sum_{i=3}^{I}(\theta_{ir} - 2\theta_{(i-1)r} + \theta_{(i-2)r})^2\right).
$$

Assuming joint age effects, the full conditional of κ_θ is given by

$$
\begin{aligned}
f(\kappa_\theta|\cdot) &\propto f(\boldsymbol{\theta}|\kappa_\theta) f(\kappa_\theta) \\
&\propto \kappa_\theta^{\frac{I-2}{2}} \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{(i-1)} + \theta_{i-2})^2\right) \cdot \kappa_\theta^{a_\theta - 1} \exp(-b_\theta \kappa_\theta) \\
&= \kappa_\theta^{a_\theta + \frac{I-2}{2} - 1} \exp\left(-\kappa_\theta\left(b_\theta + \frac{1}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right)\right)
\end{aligned}
$$

which is gamma distributed with

$$
\kappa_\theta|\cdot \sim \mathrm{Ga}\left(a_\theta + \frac{I-2}{2}, b_\theta + \frac{1}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right).
$$

Using RW1 priors the derivation of the full conditional distributions is similar.

2 Correlated multivariate age-period-cohort models

We first present the full conditional distributions for the components of the latent field. Subsequently, we present those of the precision and correlation parameters.

2.1 Latent field

Consider the case of stratum-specific effects. Using a correlated random walk for the age effects $\tilde{\boldsymbol{\theta}} = (\theta_{11}, \ldots, \theta_{I1}, \theta_{12}, \ldots, \theta_{I2}, \ldots, \theta_{IR})^\top$ the full conditional is given by

$$
\begin{aligned}
f(\tilde{\boldsymbol{\theta}}|\cdot) &\propto f(\boldsymbol{\xi}|\boldsymbol{\eta}, \kappa_z, \mathbf{C}_z) f(\tilde{\boldsymbol{\theta}}|\kappa_\theta, \mathbf{C}_\theta) \\
&\propto \prod_{i=1}^{I}\prod_{j=1}^{J} \exp\left(-\frac{1}{2}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})^\top \{\kappa_z \mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})\right) \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top \{\mathbf{C}_\theta^{-1} \otimes \mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}\right) \\
&= \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top \{\mathbf{C}_\theta^{-1} \otimes \mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}} - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})^\top \{\kappa_z \mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})\right) \\
&= \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top \{\mathbf{C}_\theta^{-1} \otimes \mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}} - \frac{1}{2}\sum_{j=1}^{J}(\boldsymbol{\xi}_{j} - \boldsymbol{\eta}_{j})^\top \{\kappa_z \mathbf{C}_z^{-1} \otimes \mathbf{I}_{I\times I}\}(\boldsymbol{\xi}_{j} - \boldsymbol{\eta}_{j})\right) \\
&= \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top \{\mathbf{C}_\theta^{-1} \otimes \mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}} - \frac{1}{2}\sum_{j=1}^{J}(\tilde{\boldsymbol{\theta}} - \boldsymbol{m}_j^\theta)^\top \{\kappa_z \mathbf{C}_z^{-1} \otimes \mathbf{I}_{I\times I}\}(\tilde{\boldsymbol{\theta}} - \boldsymbol{m}_j^\theta)\right),
\end{aligned}
$$

where $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \ldots, \xi_{ijR})^\top$, $\boldsymbol{\xi}_j = (\xi_{1j1}, \ldots, \xi_{Ij1}, \xi_{1j2}, \ldots, \xi_{Ij2}, \ldots \xi_{IjR})^\top$, $\boldsymbol{\eta}_{ij} = (\eta_{ij1}, \ldots, \eta_{ijR})^\top$ and $\boldsymbol{\eta}_j = (\eta_{1j1}, \ldots, \eta_{Ij1}, \eta_{1j2}, \ldots, \eta_{Ij2}, \ldots \eta_{IjR})^\top$. Thus, $\tilde{\boldsymbol{\theta}}|\cdot \sim \mathcal{N}(\mathbf{E}_\theta^{-1}\boldsymbol{e}_\theta, \mathbf{E}_\theta^{-1})$ with

$$
\begin{aligned}
\mathbf{E}_\theta &= (\mathbf{C}_\theta^{-1} \otimes \mathbf{R}_\theta^{(2)} + \kappa_z J\{\mathbf{C}_z^{-1} \otimes \mathbf{I}_{I\times I}\}) \\
\boldsymbol{e}_\theta &= \{\kappa_z \mathbf{C}_z^{-1} \otimes \mathbf{I}_{I\times I}\}\left(\sum_{j=J}^{J} m_{1j1}^\theta, \ldots, \sum_{j=1}^{J} m_{Ij1}^\theta, \ldots, \sum_{j=1}^{J} m_{1jR}^\theta, \ldots, \sum_{j=1}^{J} m_{IjR}^\theta\right)^\top \\
m_{ijr}^\theta &= \xi_{ijr} - \mu_r - \varphi_{j(r)} - \psi_{k(r)}.
\end{aligned}
$$

For the period and cohort effects it follows analogously, that $\tilde{\boldsymbol{\varphi}}|\cdot \sim \mathcal{N}(\mathbf{E}_\varphi^{-1}\boldsymbol{e}_\varphi, \mathbf{E}_\varphi^{-1})$, where

$$
\begin{aligned}
\mathbf{E}_\varphi &= (\mathbf{C}_\varphi^{-1} \otimes \mathbf{R}_\varphi^{(2)} + \kappa_z I\{\mathbf{C}_z^{-1} \otimes \mathbf{I}_{J\times J}\}) \\
\boldsymbol{e}_\varphi &= \{\kappa_z \mathbf{C}_z^{-1} \otimes \mathbf{I}_{J\times J}\}\left(\sum_{i=1}^{I} m_{i11}^\varphi, \ldots, \sum_{i=1}^{I} m_{iJ1}^\varphi, \ldots, \sum_{i=1}^{I} m_{i1R}^\varphi, \ldots, \sum_{i=1}^{I} m_{iJR}^\varphi\right)^\top \\
m_{ijr}^\varphi &= \xi_{ijr} - \mu_r - \theta_{i(r)} - \psi_{k(r)}
\end{aligned}
$$

and $\tilde{\boldsymbol{\psi}}|\cdot \sim \mathcal{N}(\mathbf{E}_\psi^{-1}\boldsymbol{e}_\psi, \mathbf{E}_\psi^{-1})$, where

$$
\begin{aligned}
\mathbf{E}_\psi &= \mathbf{C}_\psi^{-1} \otimes \mathbf{R}_\psi^{(2)} + \kappa_z\{\mathbf{C}_z^{-1} \otimes \mathrm{diag}(q_1, \ldots, q_k)\} \\
q_k &= \sum_{\{ij:k(i,j)=k\}} 1 \quad \text{with } k = 1, \ldots, K
\end{aligned}
$$

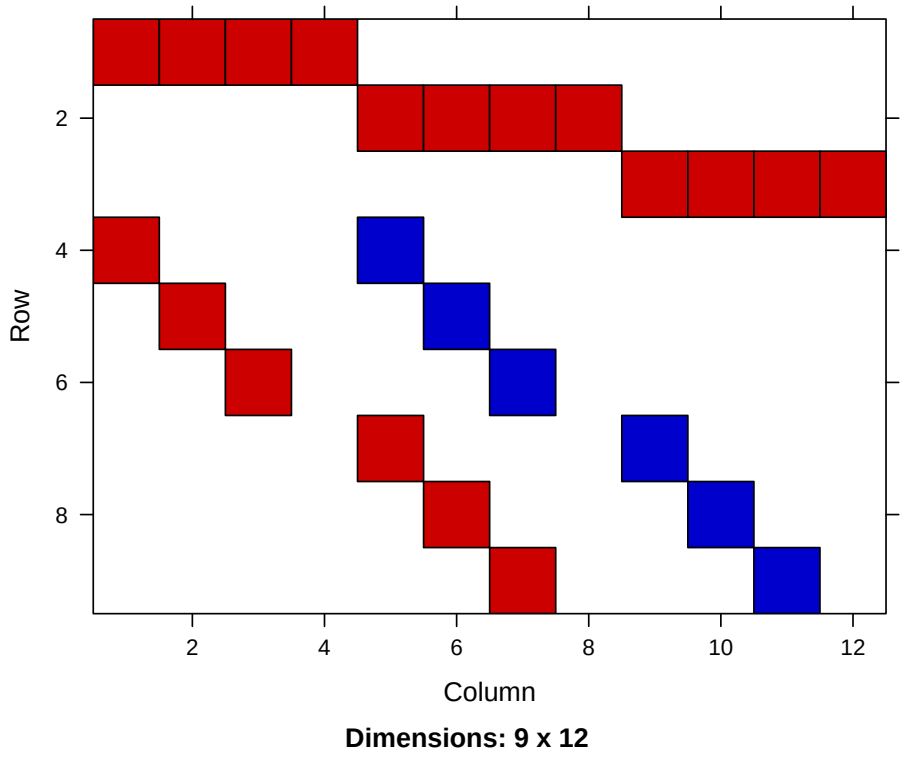

Figure 1: Image illustrating the constraint matrix $\mathbf{A}$ for sampling joint effects from the full conditional distribution of separate effects. Red squares represent the value 1 and blue squares the value -1.

$$\boldsymbol{e}_\psi = \{\kappa_z \mathbf{C}_z^{-1} \otimes \mathbf{I}_{K\times K}\}\left(\sum_{k(i,j)=1} m^\psi_{ij1}, \ldots, \sum_{k(i,j)=K} m^\psi_{ij1}, \ldots, \sum_{k(i,j)=1} m^\psi_{ijR}, \ldots, \sum_{k(i,j)=K} m^\psi_{ijR}\right)^\top$$

$$m^\psi_{ijr} = \xi_{ijr} - \mu_r - \theta_{i(r)} - \varphi_{j(r)}.$$

The full conditional distribution derived for joint effects across strata, for example $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^\top$, is not multivariate Gaussian, so that a Metropolis-Hastings step is required for sampling. However, instead we continue sampling from the full conditional distribution of the separate effects, but under $(R-1)\times(N-1)$ additional constraints. In addition to the sum-to-zero constraint imposed on each block, we constrain the effects $\boldsymbol{\theta}_r$ to be equal for all $r = 1, \ldots, R$. Thus, we sample $\tilde{\boldsymbol{\theta}}$ under a constraint $\mathbf{A}\tilde{\boldsymbol{\theta}} = \boldsymbol{\nu}$, see (Rue and Held, 2005, pages $36-40$). Figure 1 illustrates the constraint matrix $\mathbf{A}$ for $I = 4$ and $R = 3$. Red compartments represent the value 1, blue compartments the value -1 and white comparments the value 0. The vector $\boldsymbol{\nu}$ is zero-vector of length $R + (R-1)\times(N-1)$. From the obtained sampled vector $\tilde{\boldsymbol{\theta}}$, the first I elements are used as sample for $\boldsymbol{\theta}$ (then $\boldsymbol{\theta}$ repeats).

For each stratum a separate intercept μ_r with a flat prior $p(\mu_r) \propto$ const. is used, so that the

full conditional distribution for $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_R)^\top$ is given by

$$\begin{aligned}
f(\boldsymbol{\mu}|.) &\propto \prod_{i=1}^{I}\prod_{j=1}^{J} f(\boldsymbol{\xi}_{ij}|\boldsymbol{\eta}_{ij}, \kappa_z, \mathbf{C}_z) f(\boldsymbol{\mu}) \\
&\propto \prod_{i=1}^{I}\prod_{j=1}^{J} \exp\left(-\frac{1}{2}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})^\top \{\kappa_z \mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})\right) \\
&= \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\left[(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})^\top \{\kappa_z \mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})\right]\right) \\
&= \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\left[(\boldsymbol{\mu} - (\boldsymbol{\xi}_{ij} - \boldsymbol{\theta}_i - \boldsymbol{\varphi}_j - \boldsymbol{\psi}_k))^\top \{\kappa_z \mathbf{C}_z^{-1}\}(\boldsymbol{\mu} - (\boldsymbol{\xi}_{ij} - \boldsymbol{\theta}_i - \boldsymbol{\varphi}_j - \boldsymbol{\psi}_k))\right]\right),
\end{aligned}$$

so that $\boldsymbol{\mu}|\cdot \sim \mathcal{N}(\mathbf{E}_\mu^{-1}\boldsymbol{e}_\mu, \mathbf{e}_\mu^{-1})$ with

$$\begin{aligned}
\mathbf{E}_\mu &= I \cdot J \cdot \kappa_z \mathbf{C}_z^{-1} \\
\boldsymbol{e}_\mu &= \kappa_z \mathbf{C}_z^{-1}\left(\sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij1}^{\mu}, \ldots, \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ijR}^{\mu}\right)^\top \\
m_{ijr}^{\mu} &= \xi_{ijr} - \theta_{i(r)} - \varphi_{j(r)} - \psi_{k(r)}.
\end{aligned}$$

2.2 Hyperparameters

The full conditional distribution for the precision of the linear predictor/overdispersion is given by:

$$\begin{aligned}
f(\kappa_z|\cdot) &\propto f(\boldsymbol{\xi}|\boldsymbol{\eta}, \kappa_z, \mathbf{C}_z) f(\kappa_z) \\
&\propto \prod_{i=1}^{I}\prod_{j=1}^{J} |\kappa_z \mathbf{C}_z^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})^\top \{\kappa_z \mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij} - \boldsymbol{\eta}_{ij})\right) \kappa_z^{a_z - 1} \exp(-b_z \kappa_z) \\
&\propto \kappa_z^{a_z + \frac{RIJ}{2} - 1} \exp\left(-\frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{\eta})^\top \{\kappa_z \mathbf{C}_z^{-1} \otimes \mathbf{I}_{IJ \times IJ}\}(\boldsymbol{\xi} - \boldsymbol{\eta}) - b_z \kappa_z\right) \\
&= \kappa_z^{a_z + \frac{RIJ}{2} - 1} \exp\left(-\kappa_z\left(b_z + \frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{\eta})^\top \{\mathbf{C}_z^{-1} \otimes \mathbf{I}_{IJ \times IJ}\}(\boldsymbol{\xi} - \boldsymbol{\eta})\right)\right)
\end{aligned}$$

$\Rightarrow$

$$\kappa_z|\cdot \sim \text{Ga}\left(a_z + \frac{RIJ}{2}, b_z + \frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{\eta})^\top \{\mathbf{C}_z^{-1} \otimes \mathbf{I}_{IJ \times IJ}\}(\boldsymbol{\xi} - \boldsymbol{\eta})\right).$$

The full conditional distribution for precision κ_θ, if the age effects are common across strata, is

given by:

$$\begin{aligned} f(\kappa_\theta|\cdot) &\propto f(\boldsymbol{\theta}|\kappa_\theta)f(\kappa_\theta) \\ &\propto \kappa_\theta^{\frac{I-2}{2}} \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right)\kappa_\theta^{a_\theta-1}\exp(-b_\theta\kappa_\theta) \\ &= \kappa_\theta^{a_\theta+\frac{I-2}{2}-1}\exp\left(-\kappa_\theta\left(b_\theta + \frac{1}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right)\right) \end{aligned}$$

$\Rightarrow$

$$\kappa_\theta|\cdot \sim \text{Ga}\left(a_\theta + \frac{(I-2)}{2}, b_\theta + \frac{1}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right).$$

Analogously, the full conditional for the precision of the period effects would be

$$\kappa_\varphi|\cdot \sim \text{Ga}\left(a_\varphi + \frac{(J-2)}{2}, b_\varphi + \frac{1}{2}\sum_{j=3}^{J}(\varphi_j - 2\varphi_{j-1} + \varphi_{j-2})^2\right)$$

and

$$\kappa_\psi|\cdot \sim \text{Ga}\left(a_\psi + \frac{(K-2)}{2}, b_\psi + \frac{1}{2}\sum_{k=3}^{K}(\psi_k - 2\psi_{k-1} + \psi_{k-2})^2\right)$$

for the cohort effects.

In the case of stratum-specific age effects, the full conditional for the precision κ_θ follows as:

$$\begin{aligned} f(\kappa_\theta|\cdot) &= f(\tilde{\boldsymbol{\theta}}|\kappa_\theta, \mathbf{C}_\varphi)f(\kappa_\theta) \\ &\propto (|\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}|^\star)^{\frac{1}{2}}\exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}\right) \\ &\quad\times\kappa_\theta^{a_\theta-1}\exp(-b_\theta\kappa_\theta) \\ &= |\mathbf{C}_\theta^{-1}|^{\frac{I-2}{2}}\cdot|\mathbf{R}_\theta^{(2)}|^{\star\frac{R}{2}}\exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}\right) \\ &\quad\times\kappa_\theta^{a_\theta-1}\exp(-b_\theta\kappa_\theta) \\ &\propto \kappa_\theta^{a_\theta+\frac{R(I-2)}{2}-1}\exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}} - b_\theta\kappa_\theta\right) \\ &= \kappa_\theta^{a_\theta+\frac{R(I-2)}{2}-1}\exp\left(-\kappa_\theta\left(b_\theta + \frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\tilde{\mathbf{R}}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}\right)\right) \end{aligned}$$

so that

$$\kappa_\theta|\cdot \sim \text{Ga}\left(a_\theta + \frac{R(I-2)}{2}, b_\theta + \frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\tilde{\mathbf{R}}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}\right),$$

where $\tilde{\mathbf{R}}_\theta^{(2)} = \kappa_\theta^{-1}\mathbf{R}_\theta^{(2)}$. Here, we used that $(|\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}|^\star)^{\frac{1}{2}} = |\mathbf{C}_\theta^{-1}|^{\frac{I-2}{2}}(|\mathbf{R}_\theta^{(2)}|^\star)^{\frac{R}{2}}$, which follows

from (Rue and Held, 2005, page 87). Note that $\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\tilde{\mathbf{R}}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}$ can be calculated as:

$$\frac{1}{2}\sum_{r=1}^{R}\sum_{i=3}^{I}(\theta_{ir}-2\theta_{(i-1)r}+\theta_{(i-2)r})^2\cdot a_\theta+$$
$$\sum_{r=1}^{R-1}\sum_{m=r+1}^{R}\sum_{i=3}^{I}(\theta_{ir}-2\theta_{(i-1)r}+\theta_{(i-2)r})\cdot(\theta_{im}-2\theta_{(i-1)m}+\theta_{(i-2)m})\cdot b_\theta$$

Analogously, it follows:

$$\kappa_\varphi|\cdot\sim\text{Ga}\left(a_\varphi+\frac{R(J-2)}{2},b_\varphi+\frac{1}{2}\tilde{\boldsymbol{\varphi}}^\top\{\mathbf{C}_\varphi^{-1}\otimes\tilde{\mathbf{R}}_\varphi^{(2)}\}\tilde{\boldsymbol{\varphi}}\right),$$

and

$$\kappa_\psi|\cdot\sim\text{Ga}\left(a_\psi+\frac{R(K-2)}{2},b_\psi+\frac{1}{2}\tilde{\boldsymbol{\psi}}^\top\{\mathbf{C}_\psi^{-1}\otimes\tilde{\mathbf{R}}_\psi^{(2)}\}\tilde{\boldsymbol{\psi}}\right).$$

Correlations

The full conditional distribution of the transformed correlation parameter $\rho_\theta^\star$ is given by

$$\begin{aligned}f(\rho_\theta^\star|\cdot)&\propto f(\tilde{\boldsymbol{\theta}}|\kappa_\theta,\mathbf{C}_\theta)f(\rho_\theta^\star)\\&=\frac{(|\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}|^\star)^{\frac{1}{2}}}{(2\pi)^{\frac{R(I-2)}{2}}}\exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}\right)\times\frac{a_{\rho_\theta}^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\exp\left(-\frac{\rho_\theta^{\star 2}}{2a_{\rho_\theta}^{-1}}\right)\\&=\frac{|\mathbf{C}_\theta^{-1}|^{\frac{I-2}{2}}\cdot|\mathbf{R}_\theta^{(2)}|^{\star\,\frac{R}{2}}}{(2\pi)^{\frac{R(I-2)}{2}}}\exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}\right)\times\frac{a_{\rho_\theta}^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\exp\left(-\frac{\rho_\theta^{\star 2}}{2a_{\rho_\theta}^{-1}}\right),\end{aligned}$$

where $\mathbf{C}_\theta^{-1}$ is parameterised in terms of $\rho_\theta^\star$. The parameter $\rho_\theta^\star$ is updated using Metropolis-Hastings steps and a random walk proposal. Let $\rho_\theta^{\star(t)}$ be the current value at iteration t.

1. Propose a new value $\rho_\theta^{\star\text{new}}=\rho_\theta^{\star(t)}+\epsilon$ with $\epsilon\sim\mathcal{N}(0,\sigma^2)$, whereby σ^2 can be chosen by the user.

2. Simulate $u\sim\text{Uniform}[0,1]$ and update

$$\rho_\theta^{\star(t+1)}=\begin{cases}\rho_\theta^{\star\text{new}} & \text{if } u\leq\frac{f(\rho_\theta^{\star\text{new}}|\cdot)}{f(\rho_\theta^{\star(t)}|\cdot)},\\ \rho_\theta^{\star(t)} & \text{else.}\end{cases}$$

The quotient $\frac{f(\rho_\theta^{\star\text{new}}|\cdot)}{f(\rho_\theta^{\star(t)}|\cdot)}$ can be simplified to:

$$\frac{f(\rho_\theta^{\star\text{new}}|\cdot)}{f(\rho_\theta^{\star(t)}|\cdot)}=\frac{|\mathbf{C}_\theta^{-1\,\text{new}}|^{\frac{I-2}{2}}\exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1\,\text{new}}\otimes\mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}-\frac{1}{2}(\rho_\theta^{\star\text{new}})^2a_{\rho_\theta}\right)}{|\mathbf{C}_\theta^{-1\,(t)}|^{\frac{I-2}{2}}\exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1\,(t)}\otimes\mathbf{R}_\theta^{(2)}\}\tilde{\boldsymbol{\theta}}-\frac{1}{2}(\rho_\theta^{\star(t)})^2a_{\rho_\theta}\right)}.$$

The variance σ^2 is chosen to achieve acceptance rates around 40%. Note that $\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top\{\mathbf{C}_\theta^{-1}\otimes\mathbf{R}^{(2)}\}\tilde{\boldsymbol{\theta}}$ can be calculated as:

$$\frac{1}{2}\sum_{r=1}^{R}\sum_{i=3}^{I}(\theta_{ir}-2\theta_{(i-1)r}+\theta_{(i-2)r})^2\cdot\kappa_\theta a_\theta+ \sum_{r=1}^{R-1}\sum_{m=r+1}^{R}\sum_{i=3}^{I}(\theta_{ir}-2\theta_{(i-1)r}+\theta_{(i-2)r})\cdot(\theta_{im}-2\theta_{(i-1)m}+\theta_{(i-2)m})\cdot\kappa_\theta b_\theta.$$

Analogous derivations follow for $\rho_\varphi^\star$ and $\rho_\psi^\star$.

The full conditional for the correlation of the overdispersion is given by

$$\begin{aligned}
f(\rho_z^\star|\cdot) &\propto f(\boldsymbol{\xi}|\boldsymbol{\eta},\kappa_z,\mathbf{C}_z)f(\rho_z^\star)\\
&\propto \prod_{i=1}^{I}\prod_{j=1}^{J}|\kappa_z\mathbf{C}_z^{-1}|^{\frac{1}{2}}\exp\left(-\frac{1}{2}(\boldsymbol{\xi}_{ij}-\boldsymbol{\eta}_{ij})^\top\{\kappa_z\mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij}-\boldsymbol{\eta}_{ij})\right)\times\frac{a_{\rho_z}^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\exp\left(-\frac{\rho_z^{\star 2}}{2a_{\rho_z}^{-1}}\right)\\
&\propto |\mathbf{C}_z^{-1}|^{\frac{IJ}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}(\boldsymbol{\xi}_{ij}-\boldsymbol{\eta}_{ij})^\top\{\kappa_z\mathbf{C}_z^{-1}\}(\boldsymbol{\xi}_{ij}-\boldsymbol{\eta}_{ij})-\frac{1}{2}\rho_z^{\star 2}a_{\rho_z}\right)\\
&= |\mathbf{C}_z^{-1}|^{\frac{IJ}{2}}\exp\left(-\frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\eta})^\top\{\kappa_z\mathbf{C}_z^{-1}\otimes\mathbf{I}_{IJ\times IJ}\}(\boldsymbol{\xi}-\boldsymbol{\eta})-\frac{1}{2}\rho_z^{\star 2}a_{\rho_z}\right).
\end{aligned}$$

The update of $\rho_z^\star$ is analogous to the update of the other correlation parameters using a Metropolis-Hastings step. Note that $\frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\eta})^\top\{\kappa_z\mathbf{C}_z^{-1}\otimes\mathbf{I}_{IJ\times IJ}\}(\boldsymbol{\xi}-\boldsymbol{\eta})$ can be calculated by

$$\frac{1}{2}\sum_{r=1}^{R}\sum_{i=1}^{I}\sum_{j=1}^{J}(\xi_{ijr}-\eta_{ijr})^2\cdot\kappa_z a_z+ \sum_{r=1}^{R-1}\sum_{m=r+1}^{R}\sum_{i=1}^{I}\sum_{j=1}^{J}(\xi_{ijr}-\eta_{ijr})\cdot(\xi_{ijm}-\eta_{ijm})\cdot\kappa_z b_z.$$

References

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation of stochastic-systems, *Statistical Science* **10**: 3–41.

Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping, *Scandinavian Journal of Statistics* **29**: 597–614.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC Press, London.

APPENDIX III

Program description

Multivariate Age-Period-Cohort Models: Program Manual

Andrea Riebler and Leonhard Held
Biostatistics Unit, Institute of Social and Preventive Medicine,
University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

Abstract

This manual describes the `mapc_cor` program for the analysis of (correlated) multivariate age-period-cohort models. First, we give a reference manual for the program and then we provide a worked out example to illustrate the usage of `mapc_cor` in detail.

Introduction

This is a short description of the program `mapc_cor` for the analysis of (correlated) multivariate age-period-cohort models, see Riebler *et al.* (2010) for a detailed model description. The program was primarily written to include correlated overdispersion parameters and can analyse all types of ordinary and correlated multivariate APC models. The program was developed under Kubuntu 9.04 on a laptop with Intel(R) Core(TM) 2 Duo T7200 processor 2.0 GHz and is written in the low-level programming language `C`. It uses the GNU Scientific Library (Galassi *et al.*, 2009), a numerical library for `C` and `C++`, and `GMRFLib` (Rue and Held, 2005, Appendix B), a library in `C` for fast and exact simulation from Gaussian Markov random fields. The components of the model are to be specified in an ini-file. Then the program is started by typing

```
./mapc_cor ini-file
```

in the terminal.

Format of the input files

There are two types of input files: the data file containing the number of persons at risk n_{ijr} and the number of cases y_{ijr} for all $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $r = 1, \ldots, R$, and optionally a file with initial values for the linear predictor.
The format of the data file is:

$$n_{ijr} \quad y_{ijr}$$

The first column contains the number of persons at risk:

$$n_{111} \quad n_{211} \quad \ldots \quad n_{I11} \quad n_{121} \quad n_{221} \quad \ldots \quad n_{I21} \quad \ldots \quad n_{IJ1} \quad n_{112} \quad \ldots \quad n_{IJR}.$$

The second column contains the observations in the same order

$$y_{111} \quad y_{211} \quad \ldots \quad y_{I11} \quad y_{121} \quad y_{221} \quad \ldots \quad y_{I21} \quad \ldots \quad y_{IJ1} \quad y_{112} \quad \ldots \quad y_{IJR}.$$

The format of the file with starting values (without offset) for the linear predictor is:

$$\xi_{ijr}$$

Thus, the file has only one column with values for the linear predictor

$$\xi_{111} \quad \xi_{211} \quad \ldots \quad \xi_{I11} \quad \xi_{121} \quad \xi_{221} \quad \ldots \quad \xi_{I21} \quad \ldots \quad \xi_{IJ1} \quad \xi_{112} \quad \ldots \quad \xi_{IJR}.$$

Structure of the `ini-file`

The `ini-file` specifies all parameters for the algorithm. It is divided in nine sections. Each section starts with a tag written in squared brackets (*[tag]*). The following sections are to be specified:

The mcmc section

This section specifies the general settings of the MCMC algorithm. It consists of the following fields:

seed: The seed used for the random number generator.

Default: 1102534

burn_in: The number of burn-in iterations.

Default: 20 000

post_burn_in: The number of iterations after the burn-in.

Default: 100 000

thinning: The thinning interval.

Default: 20

The data section

This section specifies parameters of the data to be analysed. The following fields need to be specified:

number_of_strata: The number of strata R.

number_of_age_groups: The number of age groups I. (Note: All strata must have the same number of age groups.)

number_of_periods: The number of periods J. (Note: All strata must have the same number of periods.)

periods_per_age_groups: The number of periods per age group, namely the grid-factor M. For example, $M = 1$ for the case in which age group and period have equally spaced intervals. (Note: The data of all strata must have the same intervals.)

datafile : The path to the data file which contains the number of cases y_{ijr} and the corresponding number of persons at risk for all $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $r = 1, \ldots, R$.

outputfolder: The name of the sub-directory where the results are stored.

The linear predictor section

starting_values : The path to a file that includes starting values for the linear predictor (without offset). This is an optional field, if this field is deleted the linear predictor is internally initialised by $\log(y_{ijr}/n_{ijr})$ for $y_{ijr} > 0$ and by $\log(1/n_{ijr})$ otherwise.

The intercept section

joint_with: One of the strings "age", "period" or "cohort". This field is used to update the stratum-specific intercepts jointly with a stratum-specific (!) time effect. A joint update might improve the acceptance rates of the linear predictor. If this field is missing the intercepts are separately updated.

The random walk section

order: Order of the random walk (1 or 2) used for age, period and cohort effects.

The age effects section

In this section options for the age effects are specified.

separate: A boolean variable indicating whether the age effects should be the same or should vary across strata. Strings starting with "y", "Y", "t", "T" or "1" can be used to specify true values (return 1), strings starting with "n", "N", "f", "F", "0" represent false values (return 0).

initial_prec : Starting value for the precision.

Default: 4.0

parameters_prec_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameters_prec_b: Rate parameter for the gamma prior of the precision.

Default: 0.000 05

initial_cor : Starting value for the transformed correlation.

Default: 0.0

prec_cor: Precision of the Gaussian prior for the transformed correlation.

Default: 0.2

error_sd: Standard deviation for the random walk proposal of the transformed correlation. This values should be set to achieve acceptance rates between 25% and 45%.

Default: 0.5

fixed_cor : A boolean variable indicating whether the transformed correlation should be fixed to its initial value.

Default: **no**

exclude: A boolean variable indicating whether this time-scale shall be excluded from the analysis.

Default: **no**

The period effects section

In this section options for the period effects are specified.

separate: A boolean variable indicating whether the period effects should be the same or should vary across strata. For details on specification see also *age effects* section.

initial_prec : Starting value for the precision.

Default: 4.0

parameters_prec_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameters_prec_b: Rate parameter for the gamma prior of the precision.

Default: 0.000 05

initial_cor : Starting value for the transformed correlation.

Default: 0.0

prec_cor: Precision of the Gaussian prior for the transformed correlation.

Default: 0.2

error_sd: Standard deviation for the random walk proposal of the transformed correlation. This values should be set to achieve acceptance rates between 25% and 45%.

Default: 0.5

fixed_cor : A boolean variable indicating whether the transformed correlation should be fixed to its initial value.

Default: **no**

exclude: A boolean variable indicating whether this time-scale shall be excluded from the analysis.

Default: **no**

The cohort effects section

In this section options for the cohort effects are specified.

separate: A boolean variable indicating whether the cohort effects should be the same or should vary across strata. For details on specification see also *age effects* section.

initial_prec : Starting value for the precision.

Default: 4.0

parameters_prec_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameters_prec_b: Rate parameter for the gamma prior of the precision.

Default: 0.000 05

initial_cor : Starting value for the transformed correlation.

Default: 0.0

prec_cor: Precision of the Gaussian prior for the transformed correlation.

Default: 0.2

error_sd: Standard deviation for the random walk proposal of the transformed correlation. This values should be set to achieve acceptance rates between 25% and 45%.

Default: 0.5

fixed_cor : A boolean variable indicating whether the transformed correlation should be fixed to its initial value.

Default: **no**

exclude: A boolean variable indicating whether this time-scale shall be excluded from the analysis.

Default: **no**

The overdispersion section

In this section options for the overdispersion are specified.

initial_prec : Starting value for the precision.

Default: 4.0

parameters_prec_a: Shape parameter for the gamma prior of the precision.

Default: 1.0

parameters_prec_b: Rate parameter for the gamma prior of the precision.

Default: 0.005

initial_cor : Starting value for the transformed correlation.

Default: 0.0

prec_cor: Precision of the Gaussian prior for the transformed correlation.

Default: 0.2

error_sd: Standard deviation for the random walk proposal of the transformed correlation. This values should be set to achieve acceptance rates between 25% and 45%.

Default: 0.5

fixed_cor: A boolean variable indicating whether the transformed correlation should be fixed to its initial value.

Default: **no**

Output files

All output files are stored in the specified *outputfolder*. The following output files are generated:

- `dic.dat`: Contains estimates for the posterior mean deviance, deviance of the mean, effective number or parameters and deviance information criterion (DIC) separately for each stratum.
- `trace-acc-cor.dat`: Contains acceptance rates for each of the transformed correlation parameters. The acceptance rate is zero at each iteration if time effects are common or not correlated.
- `trace-acc-xi.dat`: Contains acceptance rates for each linear predictor block $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \ldots, \xi_{ijR})^\top$, $i = 1, \ldots, I$, $j = 1, \ldots, J$. The file has $(I \times J)$ columns and $(burn\_in + post\_burn\_in)/5000$ rows. The order of the columns is :

$$\boldsymbol{\xi}_{11} \quad \boldsymbol{\xi}_{21} \quad \ldots \quad \boldsymbol{\xi}_{I1} \quad \boldsymbol{\xi}_{12} \quad \boldsymbol{\xi}_{22} \quad \ldots \quad \boldsymbol{\xi}_{I2} \quad \boldsymbol{\xi}_{13} \quad \ldots \quad \boldsymbol{\xi}_{IJ}$$

- `trace-age.dat`: Contains the samples of the age effects in $R \times I$ columns. When the effects are common the samples are in the first I columns (ordered $1, \ldots, I$), then they repeat. If the effects are stratum-specific the samples of the first I columns correspond to the first stratum, the following I columns to the second and so on.
- `trace-cohort.dat`: Contains the samples of the cohort effects in $R \times K$ columns. When the effects are common the samples are in the first K columns (ordered $1, \ldots, K$), then they repeat. If the effects are stratum-specific the samples of the first K columns correspond to the first stratum, the following K columns to the second and so on.
- `trace-cor.dat`: Contains the estimated correlation parameters on both transformed and original scale in the order: age, period, cohort and overdispersion. The first two columns correspond to the age effects, the second two to the period effects, the third two to the cohort effects and the last two to the overdispersion. Note, however that columns for age, period or cohort are not provided if the effects are not correlated.

- `trace-expected.txt`: contains the samples of the expected counts for all strata in $I \times J \times R$ columns:

$$y^{\text{exp}}_{111} \quad y^{\text{exp}}_{211} \quad \cdots \quad y^{\text{exp}}_{I11} \quad y^{\text{exp}}_{121} \quad y^{\text{exp}}_{221} \quad \cdots \quad y^{\text{exp}}_{I21} \quad \cdots \quad y^{\text{exp}}_{IJ1} \quad y^{\text{exp}}_{112} \quad \cdots \quad y^{\text{exp}}_{IJR}$$

- `trace-mu.dat`: Contains the samples of the stratum-specific intercepts in R columns.

- `trace-period.dat`: Contains the samples of the period effects in $R \times J$ columns. When the effects are common the samples are in the first J columns (ordered $1, \ldots, J$), then they repeat. If the effects are stratum-specific the samples of the first K columns correspond to the first stratum, the following J columns to the second and so on.

- `trace-prec.dat`: Contains the samples of all precision parameters in 4 columns: κ_z, κ_θ, κ_φ, κ_ψ (overdispersion, age, period, cohort).

- `trace-xi.txt`: Contains the samples of the linear predictor (without offset) for all strata in $I \times J \times R$ columns:

$$\xi_{111} \quad \xi_{211} \quad \cdots \quad \xi_{I11} \quad \xi_{121} \quad \xi_{221} \quad \cdots \quad \xi_{I21} \quad \cdots \quad \xi_{IJ1} \quad \xi_{112} \quad \cdots \quad \xi_{IJR}$$

- `trace-yrep.txt`: Contains the samples of the replicated data points y^{rep}_{ijr}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $r = 1, \ldots, R$. The values are used to calculate the mean (multivariate) ranked probability score and the mean (multivariate) Dawid-Sebastiani score ($\overline{\text{DSS}}$)

$$y^{\text{rep}}_{111} \quad y^{\text{rep}}_{211} \quad \cdots \quad y^{\text{rep}}_{I11} \quad y^{\text{rep}}_{121} \quad y^{\text{rep}}_{221} \quad \cdots \quad y^{\text{rep}}_{I21} \quad \cdots \quad y^{\text{rep}}_{IJ1} \quad y^{\text{rep}}_{112} \quad \cdots \quad y^{\text{rep}}_{IJR}$$

Example: Mortality of Danish and Norwegian women

We illustrate the usage of `mapc_cor` by re-analysing overall mortality rates of Danish and Norwegian women (Jacobsen *et al.*, 2004). Data are provided for $R = 2$ strata, $I = 17$ age groups ($0-4$, $5-9$, ..., $80-84$) and $J = 8$ periods ($1960-1964$, ..., $1995-1999$). We assume that the number of deaths is Poisson distributed with mean $n_{ijr}\lambda_{ijr}$. Assuming shared period effects, the linear predictor is defined as

$$\xi_{ijr} = \log(\lambda_{ijr}) = \mu_r + \theta_{ir} + \varphi_j + \psi_{kr} + z_{ijr},$$

with $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$ and $r = 1, \ldots, R$. Since age group and period intervals are equally spaced $K = (I-1)+J = (17-1)+8 = 24$. Here, μ_r, is the region-specific intercept, θ_{ir} the region-specific age effect, φ_j the joint period effect, ψ_{kr} the region-specific cohort effect and z_{ijr} the overdispersion parameter.
Assume we would like to analyse a model in which the age effects, cohort effects and overdispersion parameters are correlated between Danish and Norwegian women. The corresponding `ini-file`, called `dk_n_coraPc_z.ini`, is:

```
[mcmc]
seed = 1102534
burn_in = 20000
post_burn_in = 100000
thinning = 20

[data]
number_of_strata = 2
number_of_age_groups = 17
number_of_periods = 8
periods_per_age_groups = 1

datafile =  dk_n_mcmc.dat

outputfolder = ./dk_n/coraPc_z/

[linear predictor]
starting_values = initial_values_dkn_aPc.dat

[intercept]
joint_with = cohort

[random walk]
order = 2

[age effects]
separate = yes
initial_prec = 0.8
parameters_prec_a = 1.0
parameters_prec_b = 0.00005
```

```
31 initial_cor = 0.0
32 prec_cor = 0.2
33 error_sd = 1.2
34 fixed_cor = no
35 exclude = no
36
37 [period effects]
38 separate = no
39 initial_prec = 5.0
40 parameters_prec_a = 1.0
41 parameters_prec_b = 0.00005
42 initial_cor = 0.0
43 prec_cor = 0.2
44 error_sd = 0.45
45 fixed_cor = yes
46 exclude = no
47
48 [cohort effects]
49 separate = yes
50 initial_prec = 5.0
51 parameters_prec_a = 1.0
52 parameters_prec_b = 0.00005
53 initial_cor = 0.0
54 prec_cor = 0.2
55 error_sd = 0.85
56 fixed_cor = no
57 exclude = no
58
59 [overdispersion]
60 initial_prec = 5.7
61 parameters_prec_a = 1.0
62 parameters_prec_b = 0.005
63 initial_cor = 0.0
64 prec_cor = 0.2
65 error_sd = 0.35
66 fixed_cor = no
```

The first section of the `ini-file` specifies general settings for the MCMC algorithm. Here, a burn-in of 20 000 iterations (line 3), followed by 100 000 post-burn-in iterations (line 4) is defined. The ***thinning*** variable (line 5) specifies that the samples of every 20th iteration should be stored.

The number of strata, age groups and periods, as well as the grid factor, e. g. the number of periods per age group are specified in the second section (lines 8-11). Lines 13 specifies the path to the data file. The directory in which the samples will be stored is defined in line 15.

In the third section, a file with initial values for the linear predictor can be specified (line 18). In the fourth section it can be specified whether the intercept should be directly sampled (then line 21 would be removed) or whether it should be updated together with a region-specific time effect (here the cohort effects). Updating the intercept together with a time effect might be

sometimes beneficial in terms of mixing and convergence.
In the fifth section the order of the random walk for age, period and cohort effects is set. Here, a second order random walk is defined (line 24) for age, period and cohort effects.
The following three sections specify the settings of the age, period and cohort effects. We specify separate age and cohort, but joint period effects using the variable ***separate*** (lines 27, 38, 49). The initial values for the correlations are set to zero. To specify correlated age and cohort effects we set ***fixed_cor***=no in the two corresponding sections (lines 34, 56). The precision of the Gaussian prior for the transformed correlation parameters is set to 0.2 (lines 32, 54). The shape and rate parameter of the gamma prior for the precisions are set to 1.0 and 0.000 05 for all time effects. Initial values for the precisions are set using the variable ***initial_prec***.
In the last section the initial value and parameters of the gamma prior for the precision of the overdispersion are specified. In addition, the use of correlated overdispersion parameters can be specified using the variable ***fixed_cor***, which is set to "no" for our model.

References

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M. and Rossi, F. (2009). *GNU Scientific Library Reference Manual - Third Edition*, Network Theory Ltd. ISBN 0954612078.

Jacobsen, R., von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different?, *European Journal of Epidemiology* **19**: 117–121.

Riebler, A., Held, L. and Rue, H. (2010). Correlated multivariate age-period-cohort models, *Technical report*, University of Zurich.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC Press, London.